

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 1.52905 Seconds  
(without alignment)

1830.070 Million cell updates/sec

Title: US-09-806-703A-14

Perfect score: 112

Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	100.0	21	AA11896	Immunogenic conj
2	112	100.0	21	AA11896	Tetanus toxoid pro
3	112	100.0	21	AA11896	T-cell antigen TT3
4	112	100.0	21	AA11896	Broad range helper
5	112	100.0	21	AA11896	Tetanus toxin frag
6	112	100.0	21	AA11896	T-cell epitope pep
7	112	100.0	21	AA11896	Tetanus toxoid epi
8	112	100.0	21	AA11896	Tetanus P30 epitop
9	112	100.0	21	AA11896	Tetanus toxin T ce
10	112	100.0	21	AA11896	Foreign epitope P3

11	112	100.0	21	AA11896	Amino acid sequenc
12	112	100.0	21	AA11896	CDA+ T cell epitop
13	112	100.0	21	AA11896	Clostridium tetani
14	112	100.0	21	AA11896	Amino acid sequenc
15	112	100.0	21	AA11896	Universal tetanus
16	112	100.0	21	AA11896	Tetanus Toxoid uni
17	112	100.0	21	AA11896	Tetanus toxin T-ce
18	112	100.0	21	AA11896	HER-2 B cell pepti
19	112	100.0	21	AA11896	Tetanus toxoid TT9
20	112	100.0	21	AA11896	Tetanus toxoid TT
21	112	100.0	21	AA11896	Tetanus toxoid pre
22	112	100.0	21	AA11896	Tetanus toxoid 947
23	112	100.0	21	AA11896	PSMpep010 - P30 in
24	112	100.0	21	AA11896	PSMpep011 - P30 in
25	112	100.0	21	AA11896	PSMpep012 - P30 in
26	112	100.0	21	AA11896	LHRH-containing im
27	112	100.0	21	AA11896	Amyloid beta/cetan
28	112	100.0	21	AA11896	Synthetic immunoge
29	112	100.0	21	AA11896	Peptide comprising
30	112	100.0	21	AA11896	A peptide which ma
31	112	100.0	21	AA11896	Synthetic immunoge
32	112	100.0	21	AA11896	Tetanus toxoid 830
33	112	100.0	21	AA11896	Amyloid beta/cetan
34	112	100.0	21	AA11896	Tetanus toxoid epi
35	112	100.0	21	AA11896	Amyloid beta/cetan
36	112	100.0	21	AA11896	Synthetic immunoge
37	112	100.0	21	AA11896	Tetanus toxoid epi
38	112	100.0	21	AA11896	Amyloid beta/cetan
39	112	100.0	21	AA11896	Peptide comprising
40	112	100.0	21	AA11896	A peptide which ma
41	112	100.0	21	AA11896	Immunogenic branch
42	112	100.0	21	AA11896	Immunogenic branch
43	112	100.0	21	AA11896	Immunogenic branch
44	112	100.0	21	AA11896	Immunogenic branch
45	112	100.0	21	AA11896	Tetanus toxoid epi

#### ALIGNMENTS

RESULT 1	AA11896	standard; peptide; 21 AA.
ID	AA11896	
AC	AA11896	
XX		
DT	19-JUL-1991	(first entry)
XX		
DE	Immunogenic conjugate constituent peptide, TT3.	
XX		
KW	Malaria vaccine; major histocompatibility complex.	
XX		
OS	Tetanus toxin.	
XX		
Key	Location/Qualifiers	
FT	Peptide	1..14
FT		/label= active fragment (claimed)
XX		
XX	BP427347-A.	
XX	15-MAY-1991.	
XX		
PF	07-NOV-1990;	90EP-0202948.
XX		
PR	10-NOV-1989;	89IT-0022355.
XX		
PA	(ENIE) ENRICECHER SPA.	
XX		
PI	Bianchi E, Pessi A, Corradin G;	
XX		
DR	WPI; 1991-141874/20.	
XX		
PT	Synthetic peptide(s) used as universal carriers - for preparing	

PT immunogenic conjugates used as vaccines against Plasmodium  
 PT falciparum  
 XX  
 PS Claim 1, page 13, 16pp; English.  
 XX  
 CC This peptide corresponds to residues 947-967 of Tetanus toxin. It can  
 CC be used as a universal carrier for the prepn. of an immunogenic  
 CC conjugate. It is covalently bound to a peptide or polysaccharide  
 CC hapten derived from a pathogen. This conjugate can be used as  
 CC a vaccine for malaria. This peptide is recognised by different T-  
 CC helper cell clones in association with alleles of the human MHC.  
 CC It contains 2 epitopes: (a) 953-967, recognised by DR5-restricted  
 CC clones; and (b) 947-960, recognised by all other DR and DP-  
 CC restricted clones.  
 XX  
 SQ Sequence 21 AA;  
 Query Match 100.0%; Score 112; DB 12; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FNNFTVSEFWLRVPRKVSASHLE 21  
 DB 1 FNNFTVSEFWLRVPRKVSASHLE 21  
 RESULT 2  
 ID AAW06130 standard; Peptide; 21 AA.  
 XX  
 AC AAW06130;  
 XX  
 DT 07-FEB-1997 (first entry)  
 XX  
 DE Tetanus toxoid protein T-cell epitope.  
 XX  
 KW Cholesteryl ester transfer protein; CERP; antigen; vaccine;  
 KM cardiovascular disease; atherosclerosis; tetanus toxoid;  
 XX T-cell epitope.  
 OS Clostridium tetani.  
 XX  
 PN WO9634888-A1.  
 XX  
 PD 07-NOV-1996.  
 XX  
 PF 01-MAY-1996; 96WO-US06147.  
 XX  
 PR 01-MAY-1995; 95US-0432483.  
 XX  
 PA (TCEL-) T CELL SCI INC.  
 XX  
 PI Rittershaus CM, Thomas LJ;  
 DR WPI; 1996-506103/50.  
 XX  
 PT Cholesteryl ester transfer protein B cell epitope linked to T cell  
 PT epitope - used to generate vaccine to regulate CERP activity for  
 PT decreasing the risk of developing a cardiovascular disease e.g.  
 PT atherosclerosis  
 XX  
 PS Claim 11; Page 43; 72pp; English.  
 XX  
 CC A helper T-cell epitope (AAW06130) comprises amino acids 947-967  
 CC of tetanus toxoid protein. It can be utilised in novel peptide  
 CC vaccines (see also AAW06129, AAW06132) also including B-cell  
 CC epitope(s) from human or rabbit cholesteryl ester transfer  
 CC protein (CERP) to elicit an immune response against endogenous  
 CC CERP activity, thereby treating or preventing a cardiovascular  
 CC disease, such as atherosclerosis.  
 XX  
 SQ Sequence 21 AA;

Query Match 100.0%; Score 112; DB 17; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FNNFTVSEFWLRVPRKVSASHLE 21  
 DB 1 FNNFTVSEFWLRVPRKVSASHLE 21  
 RESULT 3  
 ID AAR88397 standard; Peptide; 21 AA.  
 XX  
 AC AAR88397;  
 XX  
 DT 12-JUN-1996 (first entry)  
 XX  
 DE T-cell antigen TT3 peptide.  
 XX  
 KW T-antigen; vaccine; antibody; T-cell; T-lymphocyte;  
 KM alpha-helix; coiled-coil heterodimer; core peptide; subunit.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9531480-A1.  
 XX  
 PD 23-NOV-1995.  
 XX  
 PF 18-MAY-1995; 95WO-CA00293.  
 XX  
 PR 18-MAY-1994; 94US-0245507.  
 XX  
 PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.  
 XX  
 PI Cachia PJ, Hodges RS, Houston ME, Irvin RT, Kay CM;  
 PI Zhou NE;  
 XX  
 DR WPI; 1996-010880/01.  
 XX  
 PT Hetero-dimeric polypeptide immunogen in coiled-coil configuration  
 PT with different antigens on each sub:unit - useful in vaccines and  
 PT for antibody produ.  
 XX  
 PS Claim 7; Page 62; 95pp; English.  
 XX  
 CC This T-cell antigen TT3 peptide may be attached to a core peptide  
 CC contained in one of the 2 subunits of an alpha-helical coiled-coil  
 CC heterodimer. Each core peptide is comprised of terminal and  
 CC internal AA repeat sequences. This peptide antigen is attached  
 CC to the core peptide through covalent linkages to certain AA of the  
 CC internal repeats. The 2 subunits of the heterodimer are arranged  
 CC in a stable alpha-helical coiled-coil configuration having a 1:1  
 CC stoichiometry, and the peptide antigen is disposed toward the outer  
 CC surfaces of the configuration. The heterodimer may be used as a  
 CC synthetic vaccine (optionally multivalent) or to generate  
 CC antibodies.  
 XX  
 SQ Sequence 21 AA;  
 Query Match 100.0%; Score 112; DB 17; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FNNFTVSEFWLRVPRKVSASHLE 21  
 DB 1 FNNFTVSEFWLRVPRKVSASHLE 21  
 RESULT 4  
 ID AAW46449 standard; Peptide; 21 AA.  
 XX  
 AC AAW46449;

XX 18-MAY-1998 (first entry)  
 DT Broad range helper T cell epitope from the tetanus toxoid protein.  
 DE  
 XX  
 XX Cholesteryl ester transfer protein; CERP; cholesteryl ester;  
 KM high density lipoprotein; HDL; very low density lipoprotein; VLDL;  
 KM low density lipoprotein; LDL; T cell epitope; antibody;  
 KM DNA plasmid-based vaccine; broad range helper T cell epitope;  
 KM treatment; cardiovascular disease.  
 XX  
 OS Clostridium tetani.  
 XX  
 PN WO9741227-A1.  
 XX  
 PD 06-NOV-1997.  
 XX  
 XX 01-MAY-1997; 97WO-US07294.  
 PF  
 XX 21-FEB-1997; 97US-0802967.  
 PR 01-MAY-1996; 96US-0640713.  
 XX  
 PA (TCEL-) T CELLS SCI INC.  
 XX  
 XX Thomas LJ;  
 PI  
 XX WPI; 1997-549731/50.  
 DR  
 XX DNA plasmid-based vaccine encodes CERP B cell and helper T cell  
 PT epitope(s) - used for elevating high density lipoprotein levels, and  
 PT for treating cardiovascular disease  
 PT  
 XX Disclosure; Page 44; 67pp; English.  
 XX  
 PS The present sequence represents a broad range helper T cell epitope  
 CC of the tetanus protein. It can be used in DNA plasmid-based vaccines  
 CC against cholesteryl ester transfer proteins (CERPs). CERPs mediate the  
 CC transfer of cholesteryl esters from high density lipoprotein (HDL)  
 CC to very low density lipoprotein (VLDL) and low density lipoprotein (LDL),  
 CC and vice versa. An increased CERP activity produces an atherogenic  
 CC lipoprotein profile and induces atherosclerosis. A DNA plasmid-based  
 CC vaccine comprises sequences encoding at least one B cell epitope of CERP  
 CC linked in frame with at least one segment encoding a broad range helper  
 CC T cell epitope. The vaccines can be used to elevate the ratio of  
 CC circulating HDL to circulating LDL, VLDL or total cholesterol in a human.  
 CC It can also be used for decreasing the level of endogenous CERP activity  
 CC in a human. The vaccine can be used to produce anti-CERP antibodies in  
 CC vivo and for treating cardiovascular disease.  
 CC  
 XX  
 XX Sequence 21 AA;  
 SQ  
 Query Match 100.0%; Score 112; DB 18; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FNNFTVSPMLRVPKVSASHLE 21  
 Db 1 FNNFTVSPMLRVPKVSASHLE 21  
 ID AAM67034 standard; peptide; 21 AA.  
 AC AAM67034;  
 XX  
 XX 15-DEC-1998 (first entry)  
 DT Tetanus toxin fragment (residues 947-967).  
 DE  
 XX  
 XX Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;  
 KM dendimeric poly-lysine; epitope; tumour.  
 XX

OS Clostridium tetani.  
 XX  
 PN WO9843677-A1.  
 XX  
 XX 08-OCT-1998.  
 PD  
 XX 27-MAR-1998; 98WO-EP01922.  
 PF  
 XX 27-MAR-1997; 97US-0041726.  
 PR  
 XX (INSP ) INST PASTEUR.  
 PA  
 XX Bay S, Cantacuzene D, Leclerc C, Lo-man R;  
 PI  
 XX WPI; 1998-557071/47.  
 DR  
 XX Carbohydrate peptide conjugate used as vaccine - comprises carrier  
 PT with dendimeric poly-lysine enabling multiple epitopes to be  
 PT covalently attached  
 PT  
 XX Disclosure; Page 13; 55pp; English.  
 PS  
 XX The invention relates to a new carbohydrate peptide conjugate, which  
 CC comprises a carrier with a dendimeric poly-lysine enabling multiple  
 CC epitopes to be covalently attached to it. Also claimed are: (1) an  
 CC antibody purified from biological fluid or cells of organisms  
 CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis  
 CC kit comprising antigen-specific antibodies elicited by immunisation with  
 CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and  
 CC diagnosis kit are used to provide pharmaceutical compositions and  
 CC vaccines against tumours. These can be used to support an immune response  
 CC against viral infections caused by hepatitis virus, HIV or cytomegalo  
 CC virus. They can be used to enhance immune responses, especially B- and T-  
 CC cell responses, of humans and animals against bacterial infections. The  
 CC carbohydrate peptide conjugate stimulates the antibody and T-cell  
 CC response without stimulating undesired immune responses. The composition  
 CC is capable of increasing the survival of tumour bearing humans and  
 CC animals. The present sequence corresponds to residues 947-967 of tetanus  
 CC toxin. The synthetic peptide corresponding to this sequence may be used  
 CC as an epitope in a carbohydrate peptide conjugate.  
 CC  
 XX  
 XX Sequence 21 AA;  
 SQ  
 Query Match 100.0%; Score 112; DB 19; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FNNFTVSPMLRVPKVSASHLE 21  
 Db 1 FNNFTVSPMLRVPKVSASHLE 21  
 ID AAM67579 standard; peptide; 21 AA.  
 AC AAM67579;  
 XX  
 XX 02-MAR-1999 (first entry)  
 DT T-cell epitope peptide #5 for chimeric fimbria/T-cell epitope peptide.  
 DE  
 XX  
 XX Chimeric, non-tyrable Haemophilus influenzae; fimbria; T-cell epitope;  
 KM immunogenic composition; immune response.  
 KM  
 XX Synthetic.  
 OS  
 XX US5843464-A.  
 PN  
 XX 01-DEC-1998.  
 PD  
 XX 02-JUN-1995; 95US-0460502.  
 PF  
 XX

```

PR 02-JUN-1995; 95US-0460502.
XX
XX (OHIS ) UNIV OHIO STATE.
XX
XX Bakaletz LO, Kaumaya PTP;
PI
XX WPI; 1999-044514/04.
XX
XX Synthetic chimeric fimbria peptide - useful for vaccination against
PT non-typable Haemophilus influenzae
XX
XX Disclosure; Column 4; 16pp; English.
XX
XX The invention relates to the manufacture of a synthetic chimeric peptide
CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via
CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is
CC used in immunogenic compositions which induce an immune response against
CC non-typable Haemophilus influenzae. This sequence represents an example
CC of a T-cell epitope peptide used to generate the chimeric peptide.
XX
XX Sequence 21 AA:
SQ
Query Match 100.0%; Score 112; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 3,9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNNFTVSFWLRVPRVSASHLE 21
DB 1 FNNFTVSFWLRVPRVSASHLE 21

RESULT 7
AAW73222
ID AAW73222 standard; Protein; 21 AA.
XX
XX AAW73222;
AC
XX 25-JAN-1999 (first entry)
DT
XX
XX Tetanus toxoid epitope.
DE
XX
XX Multispecific single chain antibody; antibody H22; tumour cell; therapy;
KM antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
XX epidermal growth factor receptor; breast cancer; ovarian cancer.
XX
XX Synthetic.
OS
XX
XX USS837243-A.
PN
XX
XX 17-NOV-1998.
PD
XX
XX 07-JUN-1996; 96US-0661052.
PF
XX
XX 07-JUN-1996; 96US-0661052.
PR
XX 07-JUN-1995; 95US-0484172.
XX
XX (MEDA-) MEDAREX INC.
PA
XX
XX Deo YM, Goldstein J, Graziano R, Somasundaram C;
PI
XX WPI; 1999-023374/02.
XX
XX Specific killing of tumour cells - using a multi-specific molecule
PT comprising an anti-Fc receptor antibody and a portion which binds to
PT a target cell
XX
XX Example 7; Column 27; 57pp; English.
XX
XX This sequence represents a tetanus toxoid epitope and is recognised
CC by the multispecific single chain antibody designated H22. The
CC antibody can be used in the method of the invention for inducing
CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
CC which is characterised by overexpression of HER 2/neu or epidermal growth

```

```

CC factor receptor (EGFR), comprises contacting the tumour cell with a
CC multispecific protein molecule (preferably a single chain antibody)
CC comprising: (a) an anti-Fc receptor antibody or an antigen binding
CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which
CC binds to EGFR. The method can be used for treating cancers especially
CC breast cancer or ovarian cancer. The multispecific antibody can also
CC be administered prophylactically to vaccinate a subject against infection
CC by a target cell.
XX
XX Sequence 21 AA:
SQ
Query Match 100.0%; Score 112; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 3,9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNNFTVSFWLRVPRVSASHLE 21
DB 1 FNNFTVSFWLRVPRVSASHLE 21

RESULT 8
AAB45512
ID AAB45512 standard; Protein; 21 AA.
XX
XX AAB45512;
AC
XX 26-FEB-2001 (first entry)
DT
XX
XX Tetanus p30 epitope SEQ ID NO: 24.
DE
XX
XX Asthma; IL-5; interleukin-5; allergy; cytokine; helminthic infection;
KM cancer; eosinophilia; vaccine; allergic rhinitis.
XX
XX Clostridium tetani.
OS
XX
XX WO200065058-A1.
PN
XX
XX 02-NOV-2000.
PD
XX
XX 19-APR-2000; 2000WO-DK00205.
PF
XX
XX 23-APR-1999; 99DK-0000552.
PR
XX 06-MAY-1999; 99US-0132811.
XX
XX (MEBT-) M & E BIOTECH AS.
PA
XX
XX Klyner S;
PI
XX
XX WPI; 2000-672791/65.
DR
XX
XX Down-regulating interleukin 5 (IL-5) activity in humans by
PT administering IL-5 and/or an IL-5 analogue, useful in the treatment,
PT prophylaxis or amelioration of asthma or other chronic allergic
PT conditions -
XX
XX Example 6; Page 137; 172pp; English.
XX
XX The present invention is concerned with methods of treating asthma,
CC eosinophilia, allergic rhinitis and other allergic diseases. These
CC involve the use of interleukin-5 (IL-5) analogues and modified IL-5
CC proteins and their coding sequences to down-regulate IL-5 activity and
CC thus reduce eosinophil numbers. The allergic diseases may be treated
CC using autovaccines, nucleic acid vaccines or live vaccines. In addition,
CC it is possible that they may be used in the treatment of cancer and
CC helminthic infections.
XX
XX Sequence 21 AA:
SQ
Query Match 100.0%; Score 112; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 3,9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNNFTVSFWLRVPRVSASHLE 21

```



Db 1 FNNFTVSFWLVRPKVSASHLE 21

## RESULT 9

ID AAY99876 standard; Protein: 21 AA.

AC AAY99876;

DT 03-OCT-2000 (first entry)

DE Tetanus toxin T cell epitope helper peptide P30.

KW Human; MAGE-10; tumour rejection antigen precursor; bladder cancer;  
KW prostate cancer; lung cancer; cancer detection; oesophageal cancer;  
KW head and neck cancer; melanoma; myeloma; sarcoma; immunogen;  
KW tetanus toxin.

XX Homo sapiens.

PN WO200026407-A1.

PD 11-MAY-2000.

PF 15-OCT-1999; 99WO-US24258.

PR 30-OCT-1998; 98US-0183714.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Boon-Falleur T, Brasseur F, Rimoldi D, Deplaen E;

DR WPI; 2000-451624/39.

PT Determining presence of cancer in samples, especially useful for  
PT detecting bladder, prostate and lung cancer comprises assaying sample  
PT for expression of tumor rejection antigen precursor MAGE-10 -

PS Example 12; Page 14; 26pp; English.

XX The present sequence is a tetanus toxin T cell epitope known as  
CC Helper peptide P30. Hybrids of this peptide and an immunogenic peptide  
CC derived from tumour rejection antigen precursor MAGE-10 were used to  
CC generate polyclonal antiserum against MAGE-10. MAGE-10 binding monoclonal  
CC antibodies can be used to detect MAGE-10 expression. A correlation  
CC between MAGE-10 expression and cancer has been discovered and thus by  
CC determining the presence of MAGE-10, the presence of cancer can be  
CC determined. MAGE-10 expression can be detected using an immunoassay, an  
CC oligonucleotide hybridisation assay or via other standard techniques.  
CC This method is especially useful for determining the presence of  
CC bladder, oesophageal, head and neck, prostate or lung cancer, or  
CC melanoma, myeloma or sarcoma.

SO Sequence 21 AA;

Query Match 100.0%; Score 112; DB 21; Length 21;

Best Local Similarity 100.0%; Pred. No. 3.9e-12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLVRPKVSASHLE 21

Db 1 FNNFTVSFWLVRPKVSASHLE 21

## RESULT 10

ID AAY92626 standard; Protein: 21 AA.

AC AAY92626;

DT 10-AUG-2000 (first entry)

DE Foreign epitope P30.

KW Foreign epitope; P2; prostate specific membrane antigen; PSM; Her2;  
KW Heregulin 2; Fibroblast growth factor 8b; FGFB; vaccination;  
KW cytotoxic T-lymphocyte immunity; self-protein; cancer; breast cancer;  
KW prostate cancer; cell-associated peptide antigen.

XX Clostridium tetani.

PN WO200020027-A2.

PD 13-APR-2000.

PF 05-OCT-1999; 99WO-DK00525.

PR 05-OCT-1998; 98DK-0001261.

PR 20-OCT-1998; 98US-0105011.

PA (MEBI-) M & B BIOTECH AS.

PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;

DR N-PSDB; AAA09461.

DR WPI; 2000-349917/30.

PT Inducing immune responses to weakly immunogenic, tumor associated

PT peptide antigens for the treatment of breast and prostate cancer

PS Example 1; Page 214; 220pp; English.

XX The claims detail a method for inducing immune responses against weakly  
CC immunogenic cell-associated peptide antigens (PA) such as those  
CC associated with cancers (i.e. self-proteins), for example, human  
CC prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or  
CC fibroblast growth factor 8b (FGFB). The method comprises effecting  
CC simultaneous presentation by antigen producing cells (APCs) of the  
CC animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte)  
CC group derived from the PA and/or at least 1 B-cell group derived from the  
CC cell-associated PA; and (2) at least 1 first T helper cell group which is  
CC foreign to the animal. Analogues of human PSM, human Her2 and  
CC human/murine FGFB comprising a substantial part of all known and  
CC predicted CTL and B-cell epitopes of the respective PA and including at  
CC least one foreign T helper epitope (e.g. P2 and/or P30) are also claimed.  
CC The method is used to treat prostate, prostate/breast or breast cancer  
CC when the PA is human PSM, FGFB and Her2, respectively.

SO Sequence 21 AA;

Query Match 100.0%; Score 112; DB 21; Length 21;

Best Local Similarity 100.0%; Pred. No. 3.9e-12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLVRPKVSASHLE 21

Db 1 FNNFTVSFWLVRPKVSASHLE 21

## RESULT 11

ID AAY84428 standard; Peptide; 21 AA.

AC AAY84428;

DT 25-JUL-2000 (first entry)

DE Amino acid sequence of the tetanus toxoid P30 epitope.

KW Osteoprotegerin ligand; OPG; osteoprotegerin; osteoclastogenesis;

KW tumour necrosis factor receptor; type II transmembrane protein;

KW osteoclast differentiation; CSF-1; osteoclast activator;

KW immune response; osteoporosis; bone resorption;

KW tetanus toxoid P30 epitope.

```

XX OS Clostridium tetani.
XX PN WO200015807-A1.
XX PD 23-MAR-2000.
XX PF 13-SEP-1999; 99WO-DK00481.
XX PR 15-SEP-1998; 98DK-0001164.
XX PR 02-OCT-1998; 98US-0102896.
XX PA (MEBI-) M & E BIOTECH AS.
XX PI Halkier T, Haaning J;
XX DR WPI; 2000-271444/23.
XX PT In vivo down-regulation of osteoprotegerin ligand (OPGL) activity used
XX PS to treat, prevent and ameliorate osteoporosis -
XX PS Example: Page 106; 110pp; English.
XX CC The present sequence represents the tetanus toxoid P30 epitope. It is
XX CC used to create a fusion protein with murine osteoprotegerin ligand
XX CC (OPGL). Osteoprotegerin is a secreted member of the tumour necrosis
XX CC factor receptor family, which blocks osteoclastogenesis in a dose
XX CC dependent manner. The OPGL protein is synthesised as a type II
XX CC transmembrane protein. The murine and human OPGL polypeptides are 87%
XX CC homologous. OPGL is a potent osteoclast differentiation factor when
XX CC combined with CSF-1. It is not capable of inducing osteoclast
XX CC differentiation in the absence of CSF-1. OPGL is also an activator of
XX CC mature osteoclasts. The specification describes a method for the in vivo
XX CC down-regulation of OPGL activity in an animal. The method comprises
XX CC using at least one OPGL polypeptide or subsequence, and/or at least one
XX CC OPGL analogue to induce an immune response in the animal. The method
XX CC and OPGL polypeptide are useful for treating, preventing and ameliorating
XX CC osteoporosis or other diseases or conditions characterised by excessive
XX CC bone resorption.
XX SQ Sequence 21 AA;
XX
XX Query Match 100.0%; Score 112; DB 21; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-12;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 FNNFTVSFWLRVPKVSASHLE 21
XX Db 1 FNNFTVSFWLRVPKVSASHLE 21
XX
XX RESULT 12
XX ID AAY49260 standard; peptide; 21 AA.
XX AC AAY49260;
XX DT 07-FEB-2000 (first entry)
XX DE CD4+ T cell epitope P30TT fragment.
XX KW Carrier protein; CD4+; T cell epitope; N6; N10; N19; immunogen;
XX KW encapsulated bacteria.
XX OS Synthetic.
XX PN WO9955730-A2.
XX PD 04-NOV-1999.
XX PR 27-APR-1999; 99WO-1B00844.
XX PR 27-APR-1998; 98GB-0008932.

```

```

XX XX (CHTR-) CHIRON SPA.
XX PA Rappuoli R, Grandi G;
XX PI WPI; 2000-023325/02.
XX DR
XX PT Carrier proteins containing CD4+ epitopes useful for protecting against
XX PT diseases caused by encapsulated bacteria -
XX PS Disclosure; Page 36; 76pp; English.
XX
XX CC The invention provides carrier proteins comprising at least 5 CD4+ T
XX CC cell epitope. The carrier protein comprises at least 1 of N6, N10 or
XX CC N19. The carrier protein can be prepared by expressing a vector
XX CC comprising a nucleic acid molecule encoding the protein, in a host cell
XX CC and recovering the expressed protein. The carrier protein can also be
XX CC produced by (a) constructing oligonucleotide molecules that encode
XX CC peptide epitopes; (b) annealing the oligonucleotides to form duplexes;
XX CC (c) introducing the duplexes into an expression vector; (d) introducing
XX CC the expression vector into a host cell; and (e) isolating the fusion
XX CC protein produced from a culture of the host cells. The carrier protein
XX CC can be used as a protective immunogen in the control of diseases caused
XX CC by encapsulated bacteria. Sequences AAY49256-266 represent CD4+ T cell
XX CC epitopes inserted in the recombinant polypeptide carrier proteins.
XX SQ Sequence 21 AA;
XX
XX Query Match 100.0%; Score 112; DB 21; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-12;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 FNNFTVSFWLRVPKVSASHLE 21
XX Db 1 FNNFTVSFWLRVPKVSASHLE 21
XX
XX RESULT 13
XX ID AAE11764 standard; peptide; 21 AA.
XX AC AAE11764;
XX DT 18-DEC-2001 (first entry)
XX DE Clostridium tetani P30 epitope.
XX KW Amyloid protein; neuroprotective; nootropic; immunostimulant; vaccine;
XX KW Alzheimer's disease; anticonvulsant; gene therapy; Pick's disease;
XX KW antidiabetic; systemic amyloidosis; maturity onset diabetes; ALS;
XX KW amyotrophic lateral sclerosis; Parkinson's disease; encephalopathy;
XX KW Huntington's disease; fronto-temporal dementia; P30 epitope.
XX OS Clostridium tetani.
XX PN WO200162284-A2.
XX PD 30-AUG-2001.
XX PF 19-FEB-2001; 2001WO-DK00113.
XX PR 21-FEB-2000; 2000DK-0000265.
XX PR 01-MAR-2000; 2000US-186295P.
XX PA (MEBI-) M & E BIOTECH AS.
XX PI Birk P, Jensen MR, Nielsen KG;
XX PR WPI; 2001-589796/66.
XX PR N-PSDB; AAD18756.
XX PT In vivo down-regulation of amyloid protein for the treatment of
XX PT Alzheimer's, comprises presenting an amyloidogenic polypeptide or its

```

PT subsequence and/or at least one analogue of the amyloidogenic  
 PT polypeptide to the immune system -  
 XX  
 PS Example 3; Page 118; 120pp; English.  
 CC The invention relates to a method for in vivo down-regulation of amyloid  
 CC protein such as beta amyloid (Abeta) in an animal, including human. The  
 CC method comprising presenting to the animal's immune system an  
 CC immunogenically effective amount of at least one amyloidogenic protein  
 CC or its subsequence and/or at least one analogue of the amyloidogenic  
 CC polypeptide. The amyloidogenic protein or its subsequence, and its  
 CC analogue is useful for the preparation of an immunogenic composition  
 CC comprising an adjuvant for down-regulating amyloid in an animal. They are  
 CC also useful in the treatment, prophylaxis or amelioration of Alzheimer's  
 CC disease or other diseases characterised by amyloid deposits. They are  
 CC also useful in the treatment of systemic amyloidosis, maturity onset  
 CC diabetes, Parkinson's disease, Huntington's disease, fronto-temporal  
 CC dementia, amyotrophic lateral sclerosis (ALS), Pick's disease and  
 CC prion-related transmissible spongiform encephalopathies. They are also  
 CC useful for inducing production of antibodies against an amyloidogenic  
 CC polypeptide. The present sequence is Clostridium tetani p30 epitope  
 CC related to the invention.  
 XX  
 SQ Sequence 21 AA;  
 CC  
 CC Query Match 100.0%; Score 112; DB 22; Length 21;  
 CC Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
 CC Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FNNFTVSFWLRVPKVSASHLE 21  
 DB 1 FNNFTVSFWLRVPKVSASHLE 21  
 RESULT 14  
 ID AAB85702 standard; peptide; 21 AA.  
 XX AAB85702;  
 AC  
 XX 29-OCT-2001 (first entry)  
 DT  
 XX Amino acid sequence of p30 epitope.  
 DE  
 XX Multivalent protein; immune response; Plasmodium vivax; parasite;  
 KW procozoicide; vaccine; malaria; recombinant; ViVac1; ViVac2.  
 XX  
 OS Plasmodium vivax.  
 XX  
 PN WO20015181-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 29-JAN-2001; 2001WO-US02937.  
 XX  
 PR 31-JAN-2000; 2000US-0179213.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Lal AA, Xiao L, Zhou Z;  
 XX  
 PI WPI; 2001-514557/56.  
 XX  
 DR New recombinant multivalent protein comprising antigenic determinants  
 XX derived from more than one stage in a life cycle of Plasmodium vivax,  
 PT useful as a vaccine for treating, preventing and reducing malarial  
 PT infection -  
 XX  
 XX Example 1; Page 25; 59pp; English.  
 PS  
 XX The invention relates to recombinant multivalent proteins (I) that  
 CC stimulate an immune response to Plasmodium vivax. (I) comprises antigenic  
 CC determinants, fragments or conservative substitutions, derived from more

CC than one stage in a life cycle of a Plasmodium vivax parasite. (I) is  
 CC useful as a vaccine for stimulating an immune response, specifically a  
 CC protective immune response that confers increased resistance to infection  
 CC by Plasmodium parasites, such as P. vivax. (I) is especially useful in  
 CC the treatment, prevention and reduction of malarial infection, as  
 CC research or diagnostic reagents for the detection of Plasmodium species  
 CC in a biological sample, and for conferring immunity against multiple  
 CC stages of the malarial parasite. The antibodies produced are useful for  
 CC the detection or measurement of antigenic epitopes derived from one or  
 CC more stages in a life cycle of a parasite, particularly P. vivax. The  
 CC vaccine comprising the recombinant proteins, is cost-effective, health-  
 CC promoting intervention for controlling, preventing or treating the  
 CC incidence of malaria. The present sequence represents the amino acid  
 CC sequence of a p30 epitope, a component of the multivalent and  
 CC multistage proteins ViVacip and ViVac2p.  
 XX  
 SQ Sequence 21 AA;  
 CC  
 CC Query Match 100.0%; Score 112; DB 22; Length 21;  
 CC Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
 CC Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FNNFTVSFWLRVPKVSASHLE 21  
 DB 1 FNNFTVSFWLRVPKVSASHLE 21  
 RESULT 15  
 ID AAB85453 standard; peptide; 21 AA.  
 XX AAB85453;  
 AC  
 XX 25-SEP-2001 (first entry)  
 DT  
 XX Universal tetanus toxin Th epitope T947-967.  
 DE  
 XX HER 2/neu; epidermal growth factor receptor; EGFR; multispecific protein;  
 KW Fc receptor; FcR; tumor cell; breast; cancer; sarcoma; carcinoma; HIV;  
 KW pathogenic; Toxoplasma gondii; candidiasis; systemic lupus; cytostatic;  
 KW immune thrombocytopenia purpura; immunosuppressive; antiviral;  
 KW antifungal; antiprotosomal; tetanus toxin.  
 XX  
 OS Clostridium tetani.  
 XX  
 PN US6270765-B1.  
 XX  
 PD 07-AUG-2001.  
 XX  
 PF 06-NOV-1996; 98US-0188082.  
 XX  
 PR 07-JUN-1996; 96US-0661052.  
 XX  
 PR 07-JUN-1995; 95US-0484172.  
 XX  
 PA (MEDA-) MEDAREX INC.  
 XX  
 PI Deo YM, Goldstein J, Graziano R, Somasundaram C;  
 XX  
 PI WPI; 2001-475189/51.  
 XX  
 DR Inducing killing of tumor cells which expresses HER 2/neu or epidermal  
 XX growth factor receptor (EGFR) by contacting the cell with multispecific  
 PT proteins comprising an anti-Fc receptor, -Her 2/neu or -EGFR antibody,  
 PT useful for treating cancer -  
 XX  
 XX Example 7; Column 29; 57pp; English.  
 PS  
 XX The invention relates to a new method for inducing killing of a tumor  
 CC cell which expresses HER 2/neu or epidermal growth factor receptor  
 CC (EGFR). The method comprises contacting the tumor cell with a  
 CC multispecific protein comprising a component, preferably an antibody,  
 CC which binds to an Fc receptor (FcR), Her 2/neu or EGFR. The method is  
 CC useful for inducing killing of a tumor cell from breast cancer, sarcoma,

CC carcinoma, or ovarian cancer. Specific multispecific proteins can also be  
 CC administered to a subject to treat or prevent other diseases or  
 CC conditions, including pathogenic infections (e.g., viral (such as HIV),  
 CC protozoan infections (such as Toxoplasma gondii), fungal infections  
 CC (such as candidiasis), and an autoimmunity (e.g., immune thrombocytopenia  
 CC purpura and systemic lupus). The present sequence represents an universal  
 CC tetanus toxin Th epitope TT947-967.  
 XX

SQ Sequence 21 AA;

Query Match 100.0%; Score 112; DB 22; Length 21;

Best Local Similarity 100.0%; Pred. No. 3.9e-12; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNNFTVSFMLRVPKVSASHLE 21  
 |||||  
 Db 1 FNNFTVSFMLRVPKVSASHLE 21

Search completed: July 22, 2003, 08:08:16  
 Job time : 2.52905 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 1.09218 Seconds  
(without alignments)  
1830.070 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq.101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	15	11 AAR06310	Tetanus toxin epit
2	74	100.0	15	18 AAM35506	Tetanus toxin epit
3	74	100.0	15	18 AAM11505	Tetanus toxin epit
4	74	100.0	15	19 AAM67033	Tetanus toxin epit
5	74	100.0	15	19 AAM71321	Tetanus toxin epit
6	74	100.0	15	20 AAY04051	Universal helper T
7	74	100.0	15	20 AAM67578	T-cell epitope pep
8	74	100.0	15	20 AAM73220	Tetanus toxoid epi
9	74	100.0	15	21 AAB45511	Tetanus P2 epitope
10	74	100.0	15	21 AAY82637	Tetanus toxoid T c

11	74	100.0	15	21 AAY92625	Foreign epitope P2
12	74	100.0	15	21 AAY64427	Amino acid sequenc
13	74	100.0	15	21 AAY70300	Clostridium tetani
14	74	100.0	15	21 AAY44763	Tetanus toxoid pro
15	74	100.0	15	22 AAE11763	Clostridium tetani
16	74	100.0	15	22 AAM99515	Vaccine related MH
17	74	100.0	15	22 AAB85701	Amino acid sequenc
18	74	100.0	15	22 AAB85451	Wild-type TT830 (c
19	74	100.0	15	22 AAB61956	Tetanus Toxoid uni
20	74	100.0	15	22 AAB20143	Tetanus toxin T-ce
21	74	100.0	15	22 AAB68636	HER-2 B cell pepti
22	74	100.0	15	22 AAB46172	Tetanus toxoid TT8
23	74	100.0	15	22 AAB49071	Tetanus toxoid TT
24	74	100.0	15	23 AAY97872	Tetanus toxin P2 (
25	74	100.0	16	20 AAM35445	T-cell stimulatory
26	74	100.0	16	20 AAY97905	Clostridium tetani
27	74	100.0	16	23 AAY93865	Clostridium tetani
28	74	100.0	16	23 AAY11413	Tetanus toxoid pre
29	74	100.0	17	23 AAY82692	Helper T cell epit
30	74	100.0	17	16 AAR82573	Tetanus toxin help
31	74	100.0	17	17 AAM05599	Tetanus toxin help
32	74	100.0	17	17 AAR88395	T-cell antigen TT2
33	74	100.0	17	21 AAY92974	HIV class II bindi
34	74	100.0	17	21 AAY80056	Pathogen derived T
35	74	100.0	17	21 AAY54539	T helper cell (Th)
36	74	100.0	17	21 AAY58768	Unidentified pepti
37	74	100.0	17	22 AAM99516	Vaccine related MH
38	74	100.0	17	22 AAG62904	Amino acid residue
39	74	100.0	17	22 AAB84435	Amino acid sequenc
40	74	100.0	17	22 AAB30941	Amino acid sequenc
41	74	100.0	17	22 AAB31029	Antigenic fragment
42	74	100.0	17	22 AAB31118	Antigenic fragment
43	74	100.0	17	22 AAB15589	Peptide 5 for pept
44	74	100.0	18	20 AAY26607	HIV-derived lipope
45	74	100.0	18	23 ABB09794	Peptide TT functio

#### ALIGNMENTS

RESULT 1  
AAR06310 standard; protein; 15 AA.  
XX  
AC AAR06310;  
XX  
DT 04-DEC-1990 (first entry)  
XX  
DE Tetanus toxin epitope.  
XX  
KM Tetanus toxin; vaccine; major histocompatibility complex; MHC;  
XX  
OS Synthetic.  
XX  
PM BP378881-A.  
XX  
PD 25-JUL-1990.  
XX  
PF 27-DEC-1989; 89BP-0203318.  
XX  
PR 16-NOV-1989; 89IT-0022409.  
PR 17-JAN-1989; 89IT-0019110.  
XX  
PA (ENIE ) ENRICERCH SPA.  
XX  
PI Pessi A, Bianchi E, Verdini AS, Corradini G;  
XX  
DR WPI; 1990-225582/30.  
PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used  
PT as universal carriers for prepn. of immunogenic conjugate(s) for  
PT use as vaccines.

XX Claim 1; Page 17; 20pp; English.  
 PS  
 CC Epitopic peptides may be used with synthetic hapten derived from  
 CC a pathogen to generate an immune response to the pathogen.  
 CC Peptides are recognised by numerous T-helper cell clones within  
 CC the context of a wide range of alleles of the human MHC.  
 CC The peptides may be used in an antimalarial vaccine inducing Ab.  
 CC response to P.falciparum.  
 CC  
 XX Sequence 15 AA;  
 SO  
 Query Match 100.0%; Score 74; DB 11; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYKANSKFIGITEL 15  
 DB 1 QYKANSKFIGITEL 15  
 RESULT 2  
 AAW35506  
 ID AAW35506 standard; peptide; 15 AA.  
 AC  
 XX AAW35506;  
 AC  
 XX 22-APR-1998 (first entry)  
 DT  
 XX Universal T-cell epitope peptide SEQ ID NO:8.  
 DE  
 XX T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;  
 KM scaffold; inhibition; metastasis; wound healing; solid phase.  
 KW  
 XX Unidentified.  
 OS  
 XX MO9738011-A1.  
 PN  
 XX 16-OCT-1997.  
 PD  
 XX 03-APR-1997; 97WO-DE00146.  
 PF  
 XX 03-APR-1996; 96DK-0000398.  
 PR  
 XX (PEPR-) PEPPERSEARCH AS.  
 PA  
 PI Heegaard PMH, Jakobsen PH;  
 XX WPI; 1997-512645/47.  
 DR  
 XX Non-dendritic peptide carrier linked to a solid phase - useful as a  
 PT diagnostic agent and as a scaffold for production of chemical  
 PT derivatives  
 PT  
 XX Example 20; Page 124; 262pp; English.  
 PS  
 CC A non-dendritic peptide carrier (A) has been developed which is coupled  
 CC through a linker to a solid phase, forming a complex of (A)-solid phase.  
 CC Where (A) comprises 10-50 amino acids capable of forming a secondary  
 CC structure in a benign buffer after liberation from the solid phase, and  
 CC further the (A)-solid phase complex comprises an immunogenic substance  
 CC and/or an immune mediator coupled on (A). The present sequence  
 CC represents a peptide used in an example from the present invention.  
 CC (A)-solid phase complex can be used as a scaffold for the production of  
 CC chemical derivatives, characterised by covalently attaching molecules at  
 CC attachment points. Alternatively (A) is used as a scaffold-peptide for  
 CC the incorporation into an Immunostimulating Complex (Iscom) resulting an  
 CC (A)-Iscom complex which is used for the chemical coupling of antigenic  
 CC substances in an aqueous solution by conjugation. (A) derivatised with  
 CC one or more peptides having fibronectin-, laminin- or vitronectin-like  
 CC binding activities can be used for the promotion of cell-attachment to  
 CC plastic surfaces, in particular to inhibit tumour growth and metastasis,  
 CC and for promotion of wound healing. Also a derivatised (A) can be used

CC for the selection of specifically-binding aptamers or as a diagnostic  
 CC agent. Such diagnostic-(A) molecules could be used to detect molecules  
 CC derived from or indicative of pregnancy or of a disease, such as an  
 CC infectious, autoimmune or cancerous disease.  
 CC  
 XX Sequence 15 AA;  
 SO  
 Query Match 100.0%; Score 74; DB 18; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYKANSKFIGITEL 15  
 DB 1 QYKANSKFIGITEL 15  
 RESULT 3  
 AAW11505  
 ID AAW11505 standard; Protein; 15 AA.  
 AC  
 XX AAW11505;  
 AC  
 XX 24-SEP-1997 (first entry)  
 DT  
 XX Tetanus toxoid universal Th epitope TT830.  
 DE  
 XX Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;  
 KM fusion protein; chimera; tetanus toxoid; helper T cell epitope;  
 KW antigen presentation; ds.  
 XX  
 OS Clostridium tetani.  
 XX  
 PN MO9640789-A1.  
 PD  
 XX 19-DEC-1996.  
 PF  
 XX 07-JUN-1996; 96WO-US09988.  
 PR  
 XX 07-JUN-1995; 95US-0484172.  
 PA (MEDA-) MEDAREX INC.  
 PI Deo YM, Goldstein J, Graziano R, Somasundaram C;  
 XX WPI; 1997-052242/05.  
 DR N-PSDB; AAT58127.  
 XX  
 PT Recombinant, multi-specific anti-Fc receptor antibody molecules -  
 PT also comprise an anti-target portion, used for the treatment of  
 PT cancer, autoimmune disease and pathogenic infection  
 PT  
 XX Example 7; Fig 24; 115pp; English.  
 PS  
 CC Synthetic DNA coding for the wild-type universal Th epitope from  
 CC tetanus toxoid, designated TT830, was fused to the 3'-end of DNA  
 CC encoding heavy chain sequences from the humanised anti-Fc gamma RI  
 CC monoclonal antibody H22. The resulting fusion protein was shown to  
 CC be significantly more efficient in antigen presentation and T cell  
 CC stimulation than the TT830 epitope alone. A similar fusion  
 CC construct was prepared coding for a mutant, antagonistic form of the  
 CC epitope (designated TT833S) fused to the anti-Fc gamma RI. The  
 CC Fab22-TT833S is at least 100 times more effective than TT833S in  
 CC inhibiting T cell activation.  
 CC  
 XX Sequence 15 AA;  
 SO  
 Query Match 100.0%; Score 74; DB 18; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYKANSKFIGITEL 15  
 DB 1 QYKANSKFIGITEL 15

```

RESULT 4
AAW67033
ID AAW67033 standard; peptide; 15 AA.
XX
AC AAW67033;
XX
DT 15-DEC-1998 (first entry)
XX
DE Tetanus toxin fragment (residues 830-844).
XX
KM Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
KM dendrimeric poly-Lysine; epitope; tumour.
XX
OS Clostridium tetani.
XX
PN MO9843677-A1.
XX
PD 08-OCT-1998.
XX
PF 27-MAR-1998; 98MO-EP01922.
XX
PR 27-MAR-1997; 97US-0041726.
XX
PA (INSP ) INST PASTEUR.
XX
PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;
XX
DR WPI; 1998-557071/47.
XX
PT Carbohydrate peptide conjugate used as vaccine - comprises carrier
PT with dendrimeric poly-Lysine enabling multiple epitopes to be
PT covalently attached
XX
PS Disclosure; Page 13; 55pp; English.
XX
XX The invention relates to a new carbohydrate peptide conjugate, which
CC comprises a carrier with a dendrimeric poly-Lysine enabling multiple
CC epitopes to be covalently attached to it. Also claimed are: (1) an
CC antibody purified from biological fluid or cells of organisms
CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
CC kit comprising antigen-specific antibodies elicited by immunisation with
CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
CC diagnosis kit are used to provide pharmaceutical compositions and
CC vaccines against tumours. These can be used to support an immune response
CC against viral infections caused by hepatitis virus, HIV or cytomegalo
CC virus. They can be used to enhance immune responses, especially B- and T-
CC cell responses, of humans and animals against bacterial infections. The
CC carbohydrate peptide conjugate stimulates the antibody and T-cell
CC response without stimulating undesired immune responses. The composition
CC is capable of increasing the survival of tumour bearing humans and
CC animals. The present sequence corresponds to residues 830-844 of tetanus
CC toxin. The synthetic peptide corresponding to this sequence may be used
CC as an epitope in a carbohydrate peptide conjugate.
XX
SQ Sequence 15 AA;

```

```

Query Match 100.0%; Score 74; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

```

```

RESULT 5
AAW71321
ID AAW71321 standard; peptide; 15 AA.
XX
AC AAW71321;
XX

```

```

DT 26-NOV-1998 (first entry)
XX
DE Universal helper T-cell epitope P2 derived from tetanus toxin.
XX
KM Liver stage; Plasmodium; Navy Yoelli Liver Stage 3 antibody; NYLS3;
KM hepatic and erythrocytic stage protein; PyHEP17; vaccine;
KM malaria parasite; teanus toxin; P2; helper T-cell epitope.
XX
OS Synthetic.
OS Clostridium tetani.
XX
PN US5814617-A.
XX
PD 29-SEP-1998.
XX
PF 07-OCT-1994; 94US-0319704.
XX
PR 07-OCT-1994; 94US-0319704.
XX
PA (USNA ) US SEC OF NAVY.
XX
PI Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;
XX
DR WPI; 1998-541794/46.
XX
XX Vaccine for protecting mammal against infection by malaria caused by
PT Plasmodium species - comprises a first nucleic acid encoding a first
PT polypeptide capable of eliciting an immune reaction against an
PT antigen expressed during the liver
XX
PS Disclosure; Column 12; 24pp; English.
XX
XX AAW71321-22 represent universal helper T-cell epitopes derived from
CC tetanus toxin. They are used to enhance host immune response to
CC vaccines. The specification describes a Plasmodium yoelli liver stage
CC 17 kDa hepatic and erythrocytic stage protein designated PyHEP17. This
CC protein elicits a response from an Ig1 monoclonal antibody designated
CC Navy Yoelli Liver Stage 3 (NYLS3). This antibody does not recognise
CC sporozoites, but does recognise P. yoelli liver stage parasites. NYLS3
CC eliminates upto 90% of liver stage parasites. The specification
CC describes a vaccine for reducing the severity or incidence of infection
CC by a malaria parasite of the genus Plasmodium. The DNA vaccine comprises
CC exon 1 and part of exon 2 of the PyHEP17 gene.
XX
SQ Sequence 15 AA;

```

```

Query Match 100.0%; Score 74; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

```

```

RESULT 6
AAV04051

```

```

ID AAV04051 standard; peptide; 15 AA.
XX
AC AAV04051;
XX

```

```

DT 04-JAN-2000 (first entry)
XX
DE T-Helper epitope from tetanus toxoid.
XX
KM Covalently reactive antigen analog; CRAA; catalytic antibody;
KM electrophilic reaction centre; phosphate; boronate; vaccine;
KM transition state analog; TSA; isostere; gp120; HIV-1; T-helper;
KM tetanus; toxoid; B-T-epitope.
XX
OS Clostridium tetani.
XX
PN MO948925-A1.

```

XX 30-SEP-1999.  
 PD 23-MAR-1999; 99WO-US06325.  
 XX 23-MAR-1998; 98US-0046373.  
 PR 23-MAR-1998; 98US-0046373.  
 XX (UYNE-) UNIV NEBRASKA.  
 PA Paul S, Gololobov G, Smith L;  
 PI WPI; 1999-591076/50.  
 DR WPI; 1999-591076/50.  
 XX New covalently reactive antigen analogs used for treating e.g.  
 PT autoimmune diseases, lymphoproliferative disorders, cancers, microbial  
 PT infections, ischemic and reperfusion injury or septic shock -  
 XX  
 PS Disclosure; Page 86; 158pp; English.  
 CC The patent discloses new covalently reactive antigen analogs (CRAA)  
 CC of formula XI-Y-E-X2, in which XI and X2 represent peptide sequences  
 CC of an epitope of a disease-associated protein, Y is a positively  
 CC charged amino acid residue, preferably Lys or Arg, and E is an  
 CC electrophilic reaction centre, preferably a phosphonate or boronate  
 CC moiety. Depending on the identity of the epitope, the CRAA may be used  
 CC to stimulate production of catalytic antibodies specific for  
 CC predetermined antigens associated with particular medical disorders.  
 CC They may also be used to permanently inactivate endogenously produced  
 CC catalytic antibodies produced in certain autoimmune diseases as well as  
 CC in certain lymphoproliferative disorders.  
 CC Amongst the specifically exemplified CRAAs is one based on residues  
 CC 421-436 of a B-cell epitope of gp120 (see AAY04046) which may be used  
 CC to counter HIV-1 infections. When used as an immunogen, preferably this  
 CC CRAA is conjugated at its N-terminal to a T-helper epitope from tetanus  
 CC toxoid. The present sequence represents the T-helper epitope and  
 CC corresponds to residues 830-844 of the toxoid.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 74; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYKANSKFIGITEL 15  
 Db 1 QYKANSKFIGITEL 15  
 RESULT 7  
 AAM67578  
 ID AAM67578 standard; peptide; 15 AA.  
 AC AAM67578;  
 XX  
 DT 02-MAR-1999 (first entry)  
 XX  
 DE T-cell epitope peptide #4 for chimeric fimbria/T-cell epitope peptide.  
 XX  
 KW Chimeric; non-typable Haemophilus influenzae; fimbria; T-cell epitope;  
 KW immunogenic composition; immune response.  
 XX  
 OS Synthetic.  
 OS  
 PN US5843464-A.  
 XX  
 PD 01-DEC-1998.  
 XX  
 PF 02-JUN-1995; 95US-0460502.  
 XX  
 PR 02-JUN-1995; 95US-0460502.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX

PI Bakaletz LO, Kaumaya PTP;  
 XX  
 DR WPI; 1999-044514/04.  
 XX  
 PT Synthetic chimeric fimbria peptide - useful for vaccination against  
 PT non-typable Haemophilus influenzae  
 XX  
 PS Disclosure; Column 4; 16pp; English.  
 XX  
 CC The invention relates to the manufacture of a synthetic chimeric peptide  
 CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via  
 CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is  
 CC used in immunogenic compositions which induce an immune response against  
 CC non-typable Haemophilus influenzae. This sequence represents an example  
 CC of a T-cell epitope peptide used to generate the chimeric peptide.  
 CC  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 74; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYKANSKFIGITEL 15  
 Db 1 QYKANSKFIGITEL 15  
 RESULT 8  
 AAM73220  
 ID AAM73220 standard; Protein; 15 AA.  
 XX  
 AC AAM73220;  
 XX  
 DT 25-JAN-1999 (first entry)  
 XX  
 DE Tetanus toxoid epitope.  
 XX  
 KW Multispecific single chain antibody; antibody H22; tumour cell; therapy;  
 KW antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;  
 KW epidermal growth factor receptor; breast cancer; ovarian cancer.  
 XX  
 OS Synthetic.  
 OS  
 PN US5637243-A.  
 XX  
 PD 17-NOV-1998.  
 XX  
 PF 07-JUN-1996; 96US-0661052.  
 XX  
 PR 07-JUN-1996; 96US-0661052.  
 PR 07-JUN-1995; 95US-0484172.  
 XX  
 PA (MEDA-) MEDAREX INC.  
 XX  
 PI Deo YM, Goldstein J, Graziano R, Somasundaram C;  
 XX  
 DR WPI; 1999-023374/02.  
 XX  
 PT Specific killing of tumour cells - using a multi-specific molecule  
 PT comprising an anti-Fc receptor antibody and a portion which binds to  
 PT a target cell  
 XX  
 PS Example 7; Column 27; 57pp; English.  
 XX  
 CC This sequence represents a tetanus toxoid epitope and is recognised  
 CC by the multispecific single chain antibody designated H22. The  
 CC antibody can be used in the method of the invention for inducing  
 CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell  
 CC which is characterised by overexpression of HER 2/neu or epidermal growth  
 CC factor receptor (EGFR), comprises contacting the tumour cell with a  
 CC multispecific protein molecule (preferably a single chain antibody)  
 CC comprising: (a) an anti-Fc receptor antibody or an antigen binding  
 CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which



CC binds to EGFR. The method can be used for treating cancers especially  
 CC breast cancer or ovarian cancer. The multispecific antibody can also  
 CC be administered prophylactically to vaccinate a subject against infection  
 CC by a target cell.  
 XX

SO Sequence 15 AA;

Query Match 100.0%; Score 74; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15  
 DB 1 QYIKANSKFIGITEL 15

# RESULT 9

AAB45511 ID AAB45511 standard; Protein; 15 AA.

AC AAB45511;

DT 26-FEB-2001 (first entry)

DE Tetanus P2 epitope SEQ ID NO: 23.

KW Asthma; IL-5; interleukin-5; allergy; cytokine; helminthic infection;  
 KW cancer; eosinophilia; vaccine; allergic rhinitis.

OS Clostridium tetani.

PN WO200065058-A1.

PD 02-NOV-2000.

PF 19-APR-2000; 2000WO-DK00205.

PR 23-APR-1999; 99DK-0000552.

PR 06-MAY-1999; 99US-0132811.

PA (MEBI-) M & E BIOTECH AS.

PI Klyser S;

DR WPI; 2000-672791/65.

XX Down-regulating interleukin 5 (IL-5) activity in humans by  
 PT administering IL-5 and/or an IL-5 analogue, useful in the treatment,  
 PT prophylaxis or amelioration of asthma or other chronic allergic  
 PT conditions -

XX Example 1; Page 137; 172pp; English.

XX The present invention is concerned with methods of treating asthma,  
 CC eosinophilia, allergic rhinitis and other allergic diseases. These  
 CC involve the use of interleukin-5 (IL-5) analogues and modified IL-5  
 CC proteins and their coding sequences to down-regulate IL-5 activity and  
 CC thus reduce eosinophil numbers. The allergic diseases may be treated  
 CC using autovaccines, nucleic acid vaccines or live vaccines. In addition,  
 CC it is possible that they may be used in the treatment of cancer and  
 CC helminthic infections.  
 XX

SO Sequence 15 AA;

Query Match 100.0%; Score 74; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15  
 DB 1 QYIKANSKFIGITEL 15

RESULT 10  
 AAY82637 ID AAY82637 standard; peptide; 15 AA.

AC AAY82637;

DT 07-AUG-2000 (first entry)

DE Tetanus toxoid T cell epitope peptide SEQ ID NO:13.

KW T cell epitope; B cell epitope; allergy; allergen; antigenic;  
 KW antiallergic; antisthmatic; antiinflammatory; dermatological;  
 KW immunosuppressive; vaccine; rhinitis; sinusitis; bronchial asthma;  
 KW atopic dermatitis; acute urticaria; chronic urticaria;  
 KW gastro-intestinal syndrome; food allergen; oro-pharyngeal syndrome;  
 KW anaphylactic reaction; drug hypersensitivity; allergic reaction.  
 XX

OS Clostridium tetani.

OS Synthetic.

PN WO200006694-A2.

PD 10-FEB-2000.

PF 20-JUL-1999; 99WO-BE00092.

PR 30-JUL-1998; 98EP-0870167.

PA (UNIO) UCB SA.

PI Saint-Remy J, Jacquemin M;

DR WPI; 2000-422470/36.

XX New compound for prevention and treatment of allergies comprises at  
 PT least one allergen antigenic determinant recognized by a B cell and at  
 PT least one antigenic determinant which does not trigger T cell  
 PT activation -

XX Example 6; Page 30; 50pp; English.

XX The present invention describes a compound (I) for the prevention and/or  
 CC treatment of allergy. The compound comprises at least one allergen  
 CC antigenic determinant (i) recognised by a B cell or an antibody secreted  
 CC by a B cell of a non-atopic individual and at least one antigenic  
 CC determinant (ii) different from the allergen that triggers T cell  
 CC activation. (I) has antiallergic, antisthmatic, antiinflammatory,  
 CC dermatological and immunosuppressive activities, and can be used in a  
 CC vaccine. (I) may be used in a pharmaceutical or cosmetic medicament to  
 CC treat and/or prevent allergies or a disease of allergic origin,  
 CC especially hypersensitivities. These include rhinitis, sinusitis,  
 CC bronchial asthma, atopic dermatitis, some forms of acute and chronic  
 CC urticaria, gastro-intestinal syndromes associated with the ingestion of  
 CC food allergens, oro-pharyngeal syndrome, anaphylactic reactions  
 CC associated with drug hypersensitivities and/or a mixture of these. The  
 CC use of (I) in the treatment of allergic conditions avoids the need for  
 CC drug treatment, which often causes undesirable side-effects. Also, prior  
 CC art drug therapies alleviate symptoms, but do not influence their  
 CC causes, however (I) actually combats the cause of an allergic reaction.  
 CC The present sequence represents a peptide, which is used in an  
 CC example from the present invention.  
 XX

SO Sequence 15 AA;

Query Match 100.0%; Score 74; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15  
 DB 1 QYIKANSKFIGITEL 15

```

RESULT 11
AA92625
ID AAY92625 standard; Protein; 15 AA.
XX
XX AAY92625;
XX
XX 10-AUG-2000 (first entry)
XX
XX Foreign epitope P2.
XX
XX Foreign epitope P2.
XX
XX Foreign epitope P2; prostate specific membrane antigen; PSM; Her2;
XX Heregulin 2; Fibroblast growth factor 8b; FGF8b; vaccination;
XX cytotoxic T-lymphocyte immunity; self-protein; cancer; breast cancer;
XX prostate cancer; cell-associated peptide antigen.
XX
XX Clostridium tetani.
XX
XX WO200020027-A2.
XX
XX 13-APR-2000.
XX
XX 05-OCT-1999; 99WO-DK00525.
XX
XX 05-OCT-1998; 98DK-0001261.
XX 20-OCT-1998; 98US-0105011.
XX
XX (MEBI-) M & E BIOTECH AS.
XX
XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
XX Gautam A, Birk P, Karlsson G;
XX
XX WPI; 2000-349917/30.
XX N-PSDB; AAA09460.
XX
XX Inducing immune responses to weakly immunogenic, tumor associated
XX peptide antigens for the treatment of breast and prostate cancer
XX
XX Example 1; Page 213; 220pp; English.
XX
XX The claims detail a method for inducing immune responses against weakly
XX immunogenic cell-associated peptide antigens (PA) such as those
XX associated with cancers (i.e. self-proteins), for example, human
XX prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or
XX fibroblast growth factor 8b (FGF8b). The method comprises effecting
XX simultaneous presentation by antigen producing cells (APCs) of the
XX animal's immune system of: (i) at least 1 CTL (cytotoxic T-lymphocyte)
XX group derived from the PA and/or at least 1 B-cell group derived from the
XX cell-associated PA; and (2) at least 1 first T helper cell group which is
XX foreign to the animal. Analogues of human PSM, human Her2 and
XX human/murine FGF8b comprising a substantial part of all known and
XX predicted CTL and B-cell epitopes of the respective PA and including at
XX least one foreign T helper epitope (e.g. P2 and/or P10) are also claimed.
XX The method is used to treat prostate, prostate/breast or breast cancer
XX when the PA is human PSM, FGF8b and Her2, respectively.
XX
XX Sequence 15 AA;
SQ
Query Match 100.0%; Score 74; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

```

```

XX
XX Amino acid sequence of the tetanus toxoid P2 epitope.
XX
XX Osteoprotegerin ligand, OPG, osteoprotegerin; osteoclastogenesis;
XX tumour necrosis factor receptor; type II transmembrane protein;
XX osteoclast differentiation; CSF-1; osteoclast activator;
XX immune response; osteoporosis; bone resorption;
XX tetanus toxoid P2 epitope.
XX
XX Clostridium tetani.
XX
XX WO200015807-A1.
XX
XX 23-MAR-2000.
XX
XX 13-SEP-1999; 99WO-DK00481.
XX
XX 15-SEP-1998; 98DK-0001164.
XX 02-OCT-1998; 98US-0102896.
XX
XX (MEBI-) M & E BIOTECH AS.
XX
XX Halkier T, Haaning J;
XX
XX WPI; 2000-271444/23.
XX
XX In vivo down-regulation of osteoprotegerin ligand (OPG) activity used
XX to treat, prevent and ameliorate osteoporosis.
XX
XX Example; Page 106; 110pp; English.
XX
XX The present sequence represents the tetanus toxoid P2 epitope. It is
XX used to create a fusion protein with murine osteoprotegerin ligand
XX (OPG). Osteoprotegerin is a secreted member of the tumour necrosis
XX factor receptor family, which blocks osteoclastogenesis in a dose
XX dependent manner. The OPG protein is synthesised as a type II
XX transmembrane protein. The murine and human OPG polypeptides are 87%
XX homologous. OPG is a potent osteoclast differentiation factor when
XX combined with CSF-1. It is not capable of inducing osteoclast
XX differentiation in the absence of CSF-1. OPG is also an activator of
XX mature osteoclasts. The specification describes a method for the in vivo
XX down-regulation of OPG activity in an animal. The method comprises
XX using at least one OPG polypeptide or subsequence, and/or at least one
XX OPG analogue to induce an immune response in the animal. The method
XX CC and OPG polypeptide are useful for treating, preventing and ameliorating
XX osteoporosis or other diseases or conditions characterised by excessive
XX bone resorption.
XX
XX Sequence 15 AA;
SQ
Query Match 100.0%; Score 74; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

```

```

RESULT 12
AA84427
ID AAY84427 standard; Peptide; 15 AA.
XX
XX AAY84427;
XX
XX 25-JUL-2000 (first entry)
XX

```

```

XX
XX Clostridium tetani tetanus toxoid T-cell epitope, P589.
XX
XX Recombinant protein; CDC/NIMH/LVAC-1, multivalent; malaria; vaccine;
XX T-cell epitope; tetanus toxoid; antigenic epitope; treatment;
XX circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;
XX liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;
XX

```

KW apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;  
 KM EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;  
 KN Pf27; antiaplastic; prevention; anti-CDC/NIIH/VAC-1 antibody.  
 XX  
 OS Clostridium tetani.  
 XX  
 PN WO200011179-A1.  
 XX  
 PD 02-MAR-2000.  
 XX  
 PP 19-AUG-1999; 99WO-US18869.  
 XX  
 PR 21-AUG-1998; 98US-0097703.  
 XX  
 PA (NAIM-) NAT INST IMMUNOLOGY.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Lai AA, Shi YP, Haenain SE;  
 DR WPI: 2000-237654/20.  
 XX  
 PT Novel recombinant protein as vaccine for treating malarial infection  
 PT comprises antigenic peptides obtained from different stages of  
 PT plasmodium falciparum life cycle -  
 XX  
 PS Claim 2; Page 17; 52pp; English.  
 XX  
 CC The present sequence is the tetanus toxoid T-cell epitope P589, derived  
 CC from Clostridium tetani. It is used in the construction of recombinant  
 CC vaccine CDC/NIIH/VAC-1, which is a multivalent, multistage malarial  
 CC vaccine. The recombinant protein comprises, melittin signal peptide,  
 CC (His)6 tag, T-cell epitope from tetanus toxoid and 21 antigenic epitopes  
 CC from circumsporozoite protein (CSP), sporozoite surface protein-2  
 CC (SSP-2), liver stage antigen-1 (LSA-1), merozoite surface protein-1  
 CC (MSP-1), MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding  
 CC antigen-175 (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete  
 CC specific antigen, Pf27. These epitopes were obtained at different stages  
 CC of the life cycle of P. falciparum. CDC/NIIH/VAC-1 vaccine has  
 CC antiparasitic activity and can be used for treatment and prevention of  
 CC malarial infections. Anti-CDC/NIIH/VAC-1 antibodies can be used for  
 CC detecting P. falciparum in biological samples.  
 XX  
 SQ Sequence 15 AA;  
 QY 1 OYIKANSKFIGITEL 15  
 DB 1 OYIKANSKFIGITEL 15  
 QY 1 OYIKANSKFIGITEL 15  
 DB 1 OYIKANSKFIGITEL 15  
 RESULT 14  
 AAY44763 standard; peptide; 15 AA.  
 AAY44763:  
 04-MAY-2000 (first entry)  
 Tetanus toxoid protein derived T-cell activating epitope P2.  
 Surface layer protein; S-layer secretion signal; antibiotic; vaccine;  
 recombinant fusion protein cleavage; enzyme; protein polymer; foodstuff;  
 antibacterial enzyme; surface glycoprotein; T-cell activating epitope;  
 P2; tetanus toxoid; IPNV; Infectious pancreatic necrosis virus.  
 Clostridium tetani.  
 WO200004170-A1.  
 27-JAN-2000.

XX	14-JUL-1999;	99WO-CA00637.
PF		
XX	14-JUL-1998;	98CA-2237704.
FR		
XX	(UYBR-) UNIV BRITISH COLUMBIA.	
PA		
XX		
PI	Smit J;	
DR	WPI; 2000-182434/16.	
XX		
PT	Cleavage of Caulobacter produced recombinant fusion proteins useful for	
PT	producing vaccine peptides	
XX		
PS	Example 2; Page 16; 33pp; English.	
XX		
CC	The patent discloses a method for cleaving a recombinant fusion protein	
CC	which is produced by Caulobacter and consists of Caulobacter surface and	
CC	layer (S-layer) protein (containing the C-terminal secretion signal) and	
CC	a target protein heterologous to Caulobacter. The cleavage of target	
CC	protein from the S-layer protein is carried out under mild acid	
CC	conditions so that cleavage occurs at appropriate-proline dipeptide site	
CC	without solubilising the protein. The cleavage is accomplished while the	
CC	fusion protein is in an insoluble aggregate form which facilitates	
CC	purification of the protein. The method is useful for producing pure	
CC	proteins including recombinant human and animal therapeutic antibiotic	
CC	and vaccine peptides, enzymes, protein polymers, and antibacterial	
CC	enzymes for foodstuffs.	
CC	The present sequence is a T-cell activating epitope p2 derived	
CC	from tetanus toxoid protein. This sequence was fused to a DNA encoding	
CC	a fragment of infectious pancreatic necrosis virus surface glycoprotein	
CC	which is a vaccine candidate. This chimeric protein was in turn fused to	
CC	DNA encoding C. crescentus S-layer secretion signal (corresponds to the	
CC	C-terminal portion of the S-layer protein from amino acid 690 onwards and	
CC	contains native Asp-Pro site) for construction of a recombinant	
CC	fusion construct which is expressed in Caulobacter and then cleaved	
CC	to recover the vaccine candidate protein.	
XX		
SQ	Sequence	15 AA;
	Query Match	100.0%; Score 74; DB 21; Length 15;
	Best Local Similarity	100.0%; Pred. No. 3.9e-07;
	Matches 15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 QYIKANSKFIGITEL 15	
Db	1 QYIKANSKFIGITEL 15	
RESULT 15		
AAE11763		
ID	AAE11763 standard; peptide; 15 AA.	
AC	AAE11763;	
XX		
DT	18-DEC-2001 (first entry)	
XX		
DE	Clostridium tetani P2 epitope.	
XX		
KW	Amyloid protein; neuroprotective; nootropic; immunostimulant; vaccine;	
KW	Alzheimer's disease; anticoagulant; gene therapy; Pick's disease;	
KW	antidiabetic; systemic amyloidosis; maturity onset diabetes; ALS;	
KW	amyotrophic lateral sclerosis; Parkinson's disease; encephalopathy;	
KW	Huntington's disease; fronto-temporal dementia; P2 epitope.	
OS	Clostridium tetani.	
XX		
PN	MO200162284-A2.	
PD		
XX	30-AUG-2001.	
XX		
PF	19-FEB-2001; 2001WO-DK00113.	
XX		

```

PR 21-FEB-2000; 2000DK-0000265.
PR 01-MAR-2000; 2000US-186295P.
XX
XX
PA (MEBI-) M & E BIOTECH AS.
XX
PI Birk P, Jensen MR, Nielsen KG;
XX
DR MPI; 2001-589796/66.
DR N-PSDB; AAD18755.
XX
XX
PT In vivo down-regulation of amyloid protein for the treatment of
PT Alzheimer's, comprises presenting an amyloidogenic polypeptide or its
PT subsequence and/or at least one analogue of the amyloidogenic
PT polypeptide to the immune system -
XX
XX
PS Example 3; Page 117; 120pp; English.
XX
CC The invention relates to a method for in vivo down-regulation of amyloid
CC protein such as beta amyloid (Abeta) in an animal, including human. The
CC method comprising presenting to the animal's immune system an
CC immunogenically effective amount of at least one amyloidogenic protein
CC or its subsequence and/or at least one analogue of the amyloidogenic
CC polypeptide. The amyloidogenic protein or its subsequence, and its
CC analogue is useful for the preparation of an immunogenic composition
CC comprising an adjuvant for down-regulating amyloid in an animal. They are
CC also useful in the treatment, prophylaxis or amelioration of Alzheimer's
CC disease or other diseases characterised by amyloid deposits. They are
CC also useful in the treatment of systemic amyloidosis, maturity onset
CC diabetes, Parkinson's disease, Huntington's disease, fronto-temporal
CC dementia, amyotrophic lateral sclerosis (ALS), Pick's disease and
CC prion-related transmissible spongiform encephalopathies. They are also
CC useful for inducing production of antibodies against an amyloidogenic
CC polypeptide. The present sequence is Clostridium tetani P2 epitope
CC related to the invention.
XX
SQ Sequence 15 AA;

```

```

Query Match 100.0%; Score 74; DB 22; Length 15;
Best local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QYIKANSKFIGITEL 15
   |||||||||||||
Db 1 QYIKANSKFIGITEL 15

```

Search completed: July 22, 2003, 08:08:15  
 Job time : 2.09218 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds  
(without alignments)  
5088.033 Million cell updates/sec

Title: SEQ4-149-163-12

Perfect score: 6810

Sequence: 1 MELAALCRWGLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6734	98.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5941	87.2	1257	1 ERB2_RAT	P06494 rattus norv
3	5931.5	87.1	1254	1 ERB2_MESAU	P06053 mesocricetu
4	3158	46.4	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3136	46.0	1210	1 EGFR_MOUSE	P01279 mus musculu
6	2975.5	43.7	1308	1 ERB4_HUMAN	Q015303 homo sapien
7	2957	43.4	1308	1 ERB4_RAT	P02956 rattus norv
8	2700.5	39.7	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2432.5	35.7	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2362.5	34.7	1339	1 ERB3_RAT	P02799 rattus norv
11	1967	28.9	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00535 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P11273 avian eryth
14	1630	23.9	540	1 ERBB_AVIER	P11337 gallus galli
15	1610	23.6	703	1 EGFR_CHICK	P24348 caenorhabdi
16	1297	19.0	1233	1 LT23_CABEL	P70424 mus musculu
17	1142.5	16.8	245	1 ERB2_MOUSE	O02466 branchiosto
18	725	10.6	1363	1 ILPR_BRALA	O9W14 mus musculu
19	708	10.4	1300	1 IPR_MOUSE	P15208 mus musculu
20	705.5	10.4	1372	1 INSR_MOUSE	P06213 homo sapien
21	705	10.4	1382	1 INSR_HUMAN	P15127 rattus norv
22	705	10.4	1383	1 INSR_RAT	P15127 rattus norv
23	702	10.3	1297	1 IRR_HUMAN	P14617 cavia porce
24	702	10.3	1607	1 MIR_LYMST	Q25410 lymnaea sta
25	699.5	10.3	1300	1 IRR_CAVPO	Q25197 hydra atten
26	669	9.8	1477	1 HTK7_HYDAT	P08069 homo sapien
27	650	9.5	1367	1 IGR_HUMAN	O60701 mus musculu
28	640	9.4	1373	1 IGR_MOUSE	P24062 rattus norv
29	636.5	9.3	1370	1 IGR_RAT	O93105 aedes aegypt
30	634	9.3	1390	1 INSR_AEDAE	P09208 drosophila
31	591.5	8.7	984	1 EPB1_CHICK	O07494 gallus galli
32	591.5	8.7	987	1 EPB4_HUMAN	P54760 homo sapien
33	591	8.7	987	1 EPB4_HUMAN	P54760 homo sapien

34	589.5	8.7	977	1 EPB2_MOUSE	Q03145 mus musculu
35	588	8.6	1114	1 RET_HUMAN	P07949 homo sapien
36	584.5	8.6	976	1 EPB2_HUMAN	P29317 homo sapien
37	577.5	8.5	987	1 EPB4_MOUSE	P54761 mus musculu
38	575.5	8.5	984	1 EPB1_RAT	P09759 rattus norv
39	575	8.4	985	1 EPB4_XENLA	O01571 xenopus lae
40	573.5	8.4	1053	1 FAK1_CHICK	O00944 gallus galli
41	571	8.4	902	1 EPBB_XENLA	Q91736 xenopus lae
42	569.5	8.4	984	1 EPB1_HUMAN	P54762 homo sapien
43	569	8.4	1068	1 FAK1_XENLA	O91738 xenopus lae
44	567	8.3	757	1 HT16_HYDAT	P53356 hydra atten
45	563	8.3	1052	1 FAK1_MOUSE	P34152 mus musculu

## ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD; PRT; 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).  
GN ERBB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=6118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Cussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Franke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139(1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=2999967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RT human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
RN [5]  
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
RN POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
RN ALPHA AND AMPHIREGULIN.  
CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

[illegible]

QY	722	RAVKYVLGSAFCTYVKGIMPGGEVNVKIPVAIKVLRENTSPANKNEIIDELVYVWAGVSP	780
Db	721	RAVKYVLGSAFCTYVKGIMPGGEVNVKIPVAIKVLRENTSPANKNEIIDELVYVWAGVSP	780
QY	781	YVSRLLIGICLSTVQLVTLQMLPBGCLLDHVENRGRLSQDLNMCWIAKMSYLEDVR	840
Db	781	YVSRLLIGICLSTVQLVTLQMLPBGCLLDHVENRGRLSQDLNMCWIAKMSYLEDVR	840
QY	841	LVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETFYHADGKVPYIKMALESILRRFT	900
Db	841	LVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETFYHADGKVPYIKMALESILRRFT	900
QY	901	HOSDWSYGVATWELMTFGCAKPYDGIIPAREIDPLEKEGRLPQPICTIDVYMWKCM	960
Db	901	HOSDWSYGVATWELMTFGCAKPYDGIIPAREIDPLEKEGRLPQPICTIDVYMWKCM	960
QY	961	IDSECRPRPRELVSEFSRWARDPQRFVYIQNEDLGPASPBLDSTFRSLLEDODMDLVDA	1020
Db	961	IDSECRPRPRELVSEFSRWARDPQRFVYIQNEDLGPASPBLDSTFRSLLEDODMDLVDA	1020
QY	1021	EYLYLPQCGFFCPDDAPFAGGWNHRRHSSSTRSGGDLITGLSESEDAPRSPPLASEG	1080
Db	1021	EYLYLPQCGFFCPDDAPFAGGWNHRRHSSSTRSGGDLITGLSESEDAPRSPPLASEG	1080
QY	1081	AGSDVPDGLGCAKGLQSLPTHPSPFLQRYSEDPVPLPSETTGVAAPLTCSPQPEVY	1140
Db	1081	AGSDVPDGLGCAKGLQSLPTHPSPFLQRYSEDPVPLPSETTGVAAPLTCSPQPEVY	1140
QY	1141	NQPDVRPQPPSPBREGPLPAPRAGATLERAKTLEPGKGVWDVAFGAVENPEYLTPO	1200
Db	1141	NQPDVRPQPPSPBREGPLPAPRAGATLERAKTLEPGKGVWDVAFGAVENPEYLTPO	1200
QY	1201	GGAAPQHPPPAFSAFNLVYWDODPPBRGAPSTFGTTPAENPEYLTGLDVPY	1255
Db	1201	GGAAPQHPPPAFSAFNLVYWDODPPBRGAPSTFGTTPAENPEYLTGLDVPY	1255
RESULT 2			
ERR2_RAT			
ID	ERR2_RAT	STANDARD;	PRT; 1257 AA.
AC	P06494;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (b1dermal growth factor		
DE	receptor-related protein).		
GN	ERRB2 OR NEU.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	(1)		
RC	SEQUENCE FROM N.A.		
RC	TISSUE=Neuroblastoma;		
RA	MEDLINE=8611662; PubMed=3945311;		
RA	Bargmann C.I., Hung M.-C., Weinberg R.A.;		
RT	"The neu oncogene encodes an epidermal growth factor receptor-related		
RT	protein.";		
RL	Nature 319:226-230(1986).		
RN	(2)		
RP	SEQUENCE OF 852-905 FROM N.A.		
RC	TISSUE=Sciatic nerve;		
RX	MEDLINE=9122560; PubMed=2025425;		
RA	Lai C., Lemke G.;		
RT	"An extended family of protein-tyrosine kinase genes differentially		
RT	expressed in the vertebrate nervous system.";		
RL	Neuron 6:691-704(1991).		
RN	(3)		
RP	STRUCTURE BY NMR OF 650-666.		
RX	MEDLINE=9215181; PubMed=1346763;		
RA	Gullik W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.		

RA	Newman R,Cumpton M.J., Stenberg M.J.E., Campbell I.D.;
RT	"Three dimensional structure of the transmembrane region of the proto-
RL	oncogenic and oncogene forms of the neu protein.";
CC	EMBO J. 11:43-48(1992).
CC	-I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC	ALTHOUGH NEUREULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC	ALPHA AND AMPHIREGULIN.
CC	-I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC	tyrosine phosphate.
CC	-I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC	THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I- PRM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC	RESIDUES (BY SIMILARITY).
CC	-I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X03362; CAA27059.1; ALT_INIT.
DR	PIR; A24562; TVRTNU.
DR	HSSP; P11362; IFKC.
DR	InterPro; IPRO000494; EGFR_L_domain.
DR	InterPro; IPRO00719; Euk_Dkinase.
DR	InterPro; IPR002174; Furin-like.
DR	InterPro; IPR001245; Tyr_pkinase.
DR	InterPro; IPR004019; YLP motif.
DR	Pfam; PF00069; pkinase; 1.
DR	Pfam; PF00757; Recept_Like; 1.
DR	Pfam; PF01030; Recep_L_domain; 2.
DR	Pfam; PF02757; YLP_2.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	SMART; SMO0261; FU; 3.
DR	SMART; SMO0219; TYKc; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW	Proto-oncogene; Disease mutation.
FT	SIGNAL 1 21 POTENTIAL.
FT	CHAIN 1 21 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT	DOMAIN 22 1257 EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM 655 677 POTENTIAL.
FT	DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 159 369 CYS-RICH.
FT	DOMAIN 473 646 CYS-RICH.
FT	DOMAIN 722 989 PROTEIN KINASE.
FT	NP_BIND 728 736 ATP (BY SIMILARITY).
FT	BINDING 755 755 ATP (BY SIMILARITY).
FT	ACT_SITE 847 847 BY SIMILARITY.
FT	DISUPEID 196 205 BY SIMILARITY.
FT	DISUPEID 200 213 BY SIMILARITY.
FT	DISUPEID 221 228 BY SIMILARITY.
FT	DISUPEID 225 236 BY SIMILARITY.
FT	DISUPEID 237 245 BY SIMILARITY.
FT	DISUPEID 241 253 BY SIMILARITY.
FT	DISUPEID 256 265 BY SIMILARITY.
FT	DISUPEID 269 296 BY SIMILARITY.
FT	DISUPEID 300 312 BY SIMILARITY.
FT	DISUPEID 316 332 BY SIMILARITY.
FT	DISUPEID 335 339 BY SIMILARITY.
FT	DISUPEID 513 522 BY SIMILARITY.
FT	DISUPEID 517 530 BY SIMILARITY.
FT	DISUPEID 533 542 BY SIMILARITY.
FT	DISUPEID 546 562 BY SIMILARITY.
FT	DISUPEID 565 578 BY SIMILARITY.

```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 87.2%; Score 5941; DB 1; Length 1257;
Best Local Similarity 87.2%; Pred. No. 6,1e-309;
Matches 1096; Conservative 51; Mismatches 108; Indels 2; Gaps 2;

QY 1 MELAALCRMGILLALLPBGAASTOVCTGDMKRLPASPEHMLBHLVQGCQVOGNTL 60
D 1 MELAAMCRMGFLALLPBGIACTOVCTGDMKRLPASPEHMLBHLVQGCQVOGNTL 60
QY 61 ELTYLPNASLSFLQDIQEVQGYVLIAHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
D 61 ELTYVPANASLSFLQDIQEVQGYVLIHNOVQVPLQRLIRVGTQLFEDKYALAVLDNR 120
QY 121 DPLNNTTPVT-GASPGGLRELQRLSTEILKGVLIQBNPOLCYODITIMKDOYIKANSK 179
D 121 DPLNNTTPVTGRLPBGRLRELQRLSTEILKGVLIQBNPOLCYODIMLVKDOVRKNNQL 180
QY 180 FIGITELNRSRACHPCSPMGKSRGWESSESDCSLTTRVYAGAGCARCKGPLPTDCHEQ 239
D 181 APVDIDNRSRACHPCSPACKDNHCWESPEDQILGTICTSCACRCKGLPTDCHEQ 240
QY 240 CAAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTESMPNPEGRTYFGASCYTAC 299
D 241 CAAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTESMPNPEGRTYFGASCYTTC 300
QY 300 PYNVLSTDVSGCTLVCPHNOEVTAEQDTCRCKSPCARVCYGLMELHREVRATSA 359
D 301 PYNVLSTEVSGCTLVCPHNOEVTAEQDTCRCKSPCARVCYGLMELHREVRATSD 360
QY 360 NIOEFACCKKIFGSLAFLEPSFDDPASNAPLOPELOVEFELEITGYLYISAMPDSL 419
D 361 NVGEFDCCKKIFGSLAFLEPSFDDPSSGIALPLARELOVEFELEITGYLYISAMPDSL 420
QY 420 PDLVSFONLOVIRGRIILHNGAYSLTLOGLSWLGRLSLELGSGLALIHNTHLCFVHT 479
D 421 RDLVSFONLRIIRGRIILHNGAYSLTLOGLSHGLRLSLELGSGLALIHNTHLCFVHT 480
QY 480 VPMDOLEFRNPHOALHTANRPEDE-CVGEGLAQHQLCARGHCKWPGPTOCNCNCGPLRQ 538
D 481 VPMDOLEFRNPHOALHTANRPEDE-CVGEGLAQHQLCARGHCKWPGPTOCNCNCGPLRQ 540
QY 539 ECVEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCFGEADQCAVACAYKDPFCVA 598
D 541 ECVEECRVMKGLPREVYSDKRLCPCHPECOPONGSVCFCGEADQCAVACAYKSSCVA 600
QY 599 RCSSGVKPDLSYPIKWFPEBEGACORCPINCHSCVDLDDKGPAPORASPLTSYSAV 658
D 601 RCSSGVKPDLSYPIKWFPEBEGACORCPINCHSCVDLDERGCPAPORASPLTSYSAV 660
QY 659 VGLILVVLGVFEGILIKRROOKIRKTYMRRLLOETELVEPLPSPGAMPPOAQRILKET 718
D 661 VGLILVVLGVFEGILIKRROOKIRKTYMRRLLOETELVEPLPSPGAMPPOAQRILKET 720
QY 719 ELRKRVYLSGAGFTYVKGIMIPDGENVKI PVALIKVRENTSPKANKEILDEAYVMAVG 778
D 721 ELRKRVYLSGAGFTYVKGIMIPDGENVKI PVALIKVRENTSPKANKEILDEAYVMAVG 780
QY 779 SPVVSRLGICLSTVQLVTLQMLPYGCLLDHVENRRLGSLDLNMCQIACKMSYLED 838

```

```

D 781 SPVVSRLGICLSTVQLVTLQMLPYGCLLDHVENRRLGSLDLNMCQIACKMSYLED 840
QY 839 VRLVHRLAARNLVKSPNNVKTDPGLARLDIDENYHADGKVPKIMWALESTILRR 898
D 841 VRLVHRLAARNLVKSPNNVKTDPGLARLDIDENYHADGKVPKIMWALESTILRR 900
QY 899 FTHQSDWVSQVYTWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYMINWKC 958
D 901 FTHQSDWVSQVYTWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYMINWKC 960
QY 959 WMIDSECRPFRELVSFSSMARDPORFVYI QNEDICPASPJDSFYRSLDEDDMDGLV 1018
D 961 WMIDSECRPFRELVSFSSMARDPORFVYI QNEDICPASPJDSFYRSLDEDDMDGLV 1020
QY 1019 DAEEYLVPOGFCPPDPAPAGGVHHRHSSSTRSGGDTLTGLFSEEEAPSPPLAS 1078
D 1021 DAEEYLVPOGFCPPDPAPAGGVHHRHSSSTRSGGDTLTGLFSEEEAPSPPLAS 1080
QY 1079 EGAGSDVFDGDLGMAKGLQSLPTHDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQE 1138
D 1081 EGAGSDVFDGDLGMAKGLQSLPTHDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQE 1140
QY 1139 YVNOQDVRPQPSREBGLPAPARPAGATLERAKTSLSPGKGVKVDVAFGAVENPEYLT 1198
D 1141 YVNOQDVRPQPSREBGLPAPARPAGATLERAKTSLSPGKGVKVDVAFGAVENPEYLT 1200
QY 1199 POGGAAPQPPAPFSPAFNLYWDDPREGRAPSTFTGTPPAENPEYLGDLV 1255
D 1201 PREGTASPPHSPAFSPAFNLYWDDPREGRAPSTFTGTPPAENPEYLGDLV 1257

RESULT 3
ERB2_MESAU
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.,
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



CC		entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC		or send an email to license@isb-sib.ch).
CC		-----
DR	EMBL; D16295; BAA03801.1; -.	
DR	HSSP; P11362; IFGK.	
DR	InterPro; IPR000494; EGFR_L_domain.	
DR	InterPro; IPR000719; Euk_Pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_Pkinase.	
DR	InterPro; IPR004019; YLP_motif.	
DR	Pfam; PF000659; pkinase; 1.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF01030; Recep_L_domain; 2.	
DR	Pfam; PF02357; YLP_2.	
DR	PfDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SMO0261; FU; 3.	
DR	SMART; SMO0219; Tyrcg; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.	
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	
KW	transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;	
KW	Proto-oncogene; Disease mutation.	
FT	SIGNAL	1
FT	CHAIN	22 1254
FT	DOMAIN	22 652
FT	TRANSMEM	653 675
FT	DOMAIN	676 1254
FT	DOMAIN	158 368
FT	DOMAIN	472 644
FT	DOMAIN	720 987
FT	NP_BIND	726 734
FT	BINDING	753 753
FT	ACT_SITE	845 845
FT	DISULFID	195 204
FT	DISULFID	199 212
FT	DISULFID	236 244
FT	DISULFID	240 252
FT	DISULFID	255 264
FT	DISULFID	268 295
FT	DISULFID	299 311
FT	DISULFID	315 331
FT	DISULFID	334 338
FT	DISULFID	511 520
FT	DISULFID	515 528
FT	DISULFID	531 540
FT	DISULFID	544 560
FT	DISULFID	563 576
FT	DISULFID	567 584
FT	DISULFID	587 596
FT	DISULFID	600 623
FT	DISULFID	626 634
FT	DISULFID	630 642
FT	MOD_RES	1139 1139
FT	MOD_RES	1247 1247
FT	CARBOHYD	68 68
FT	CARBOHYD	125 125
FT	CARBOHYD	187 187
FT	CARBOHYD	259 259
FT	CARBOHYD	530 530
FT	CARBOHYD	571 571
FT	CARBOHYD	629 629
FT	VARIANT	658 658
FT	VARIANT	659 659
SO	SEQUENCE	1254 AA; 138253 MW; 974C3791C21FBE1 CRC64;
<hr/>		
Query Match	87.1%; Score 5931.5; DB 1; Length 1254;	
Best Local Similarity	86.9%; Pred. No. 2e-108;	
Matches 1091; Conservative 59; Mismatches 104; Indels 1; Gaps 1;		
Oy	1 MELAALGRWGILLALLLPPGAASGVCTGTDMKRLPLSPETHLDMLRHLYOGCGVVGNTL	60
Ds	1 MELAALGGWGLIALLLSPGASGVTCVTGTDMKRLPLSPETHLDIVHILYOGCGVVGNTL	60

Qy	61	ELTYVPTNASHLSPLDIOIEOVGYIAIHNQVRYPLORLAIIVGTOLFEENYALAVLDNG	120
Db	61	ELTYVPAWATLSFLODIOEVGYWYLIAHSQVRHPLORLRIVGTOLFEEDYALAVLDNR	120
Qy	121	DPLNNTTYVNTASPGGLRELOLRSLTELKGVLYIQRNPOLCYODTILMMDQVYIKANSKF	180
Db	121	DPLNDVTTATRGTRPGLRELOLRSLTELKGVLYIRGNPOLCYODTYLMDVVRKINOLA	180
Qy	181	IGTELNRSRACHPCSPMCKGSRGCMGSSSEDCOSLRTVYACAGCARCKGPLPTDCCHEQC	240
Db	181	PVDIDTNRNSRACPCACAPACKCDNHCMGASPEDCQTLTGTIAPRAVPAARLPTDCCHEQC	240
Qy	241	AAGCGPRHGSCLACLFHNHSGICELYCPALVYNTDTESSMNPBERYTFGASCVTACP	300
Db	241	AAGCGPRHGSCLACLFHNHSGICELYCPALVYNTDTESSMNPBERYTFGASCVTACP	300
Qy	301	YNYLSTDVSGCTLYCPAHNOEVTAEADGTORCEKSKPCARVCYGLGEMHLEREVAUTSAN	360
Db	301	YNYLSTBVGSGCTLYCPAHNOEVTAEADGTORCEKSKPCARVCYGLGEMHLERGARAITSAN	360
Qy	361	IOEPACCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVETLEIITGYLYISAMPDSLP	420
Db	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVETLEIITGYLYISAMPDLSH	420
Qy	421	DLSPFQNLQVIRGILINNGAYSLTLQOLGISMGLSRLSGSGLAIHNTHLCFVHTY	480
Db	421	DLSPFQNLRIYRGVLIHDGAYSLALQOLGIRMLGRLSRLSGGLVLIHNTHLCFVHTY	480
Qy	481	PMDOLEFRPHOLJHTANRPDECEVGEGLACHOLCARHGCMGSGPPOCVNCSQFLRGQEC	540
Db	481	PMDOLEFRPHOLJHSGNPSBEEGGLKDFACYPLCAHGHCMGQPTQCVNCSHFLRGQEC	540
Qy	541	VEECRVLOGLPREYVYNAHRCLPCHRECOPONGSVTCFGEPAECOVCAHYKDDPFCVCAR	600
Db	541	VKECRVMKGLPREYVYNGKGLCPCHRECOQPNSTETCTGCSADQCTACPHYKDSPFVCAR	600
Qy	601	PSGVKRPDLSTYMPKPKFPDEBEGACOPPCINCTHSCVDLIDKGCFAEOPASPLTISVAVNG	660
Db	601	PSGVKRPDLSTYMPKPKFYDEEGMCOPCINCTHSCVDLIDBERGCAEOPASPATSIITAVNG	660
Qy	661	ILLVVLGVNFGVGLIYIKRROOKIRYVYMRLLQETELVEPLTPSGAMPNOAMQMLIETEL	720
Db	661	ILLVVLGVNFGVGLIYIKRROOKIRYVYMRLLQETELVEPLTPSGAMPNOAMQMLIETEL	720
Qy	721	RKVAVLGSAGFVYKGIWIPDEBNVYKPAIIVYLRNENTSPKANKEILDEAYVMAGVGP	780
Db	721	RKVAVLGSAGFVYKGIWIPDEBNVYKPAIIVYLRNENTSPKANKEILDEAYVMAGLGP	780
Qy	781	YVSRLLGICLTSTYQVLTOLMPYGCGLIDHYRENRGRGLGODLLNMCMQIAKGSYLEDR	840
Db	781	YVSRLLGICLTSTYQVLTOLMPYGCGLIDHYRENRGRGLGODLLNMCMQIAKGSYLEDR	840
Qy	841	LVHBDLAARNLVYSPHVKITDTEGLARLIDIDETVYHADGGVUPIKYMALESILPRRT	900
Db	841	LVHBDLAARNLVYSPHVKITDTEGLARLIDIDETVYHADGGVUPIKYMALESILPRRT	900
Qy	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICTIDVYMIWAKCM	960
Db	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICTIDVYMIWAKCM	960
Qy	961	IDSECRPRFRELVSSEFSRMAARDPQRFVUIONEDLGPASPLDSTFYRSLLEDDMDGLYDA	1020
Db	961	IDSECRPRFRELVSSEFSRMAARDPQRFVUIONEDLGPSSPLDSTFYRSLLEDDMDGLYDA	1020
Qy	1021	EEYVYVPOQGFPCDDPAPAGAGMWHNHRSSSTSGGDLTLGLSPEEELAPRSPPLAPSG	1080
Db	1021	EEYVYVPOQGFPCDDPAPAGAGSTAHRRHRSSSTSGGDLTLGLMEPGEPPRSPPLAPSG	1080
Qy	1081	AGSDVPDGDJGMAAGLQSLPTNHSPLQRYSEDPVPLPSETDQYVAPLTCSPQPEVY	1140
Db	1081	AGSDVFEGLGMAATGQPSISPRDLSPLQRYSEDDTLPLPSETDQYVAPLTCSPQPEVY	1140

QY 1141 NOPVRRPQPSPPREGPLPAARPAAGATLERAKTLSPGKNGVGVKDVFAFGAVENPEYLTPQ 1200  
 DB 1141 NQPEVRRQPLTEPGPLPPVRRPACATLERPKTLSPGKNGVGVKDVFTFGAVENPEYLVR 1200  
 QY 1201 GGAAPQHPPPAPFADNLYYWDQDPERGAPSTFKGTPTAENPEYLGLDVPV 1255  
 DB 1201 GGSASQPH-PPALCPAFDNLYYWDQDPSESGSPNTEGTPTAENPEYLGLDVPV 1254

RESULT 4  
 EGFR HUMAN STANDARD; PRT; 1210 AA.  
 ID EGFR\_HUMAN  
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMDB; Q9UMG5; Q92795; Q00732;  
 AC Q00688; Q9BZ52; Q9H2C9; Q9GZX1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 protein-tyrosine kinase ErbB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN 11  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 Lee J., Yarden Y., Liberman T.A., Schlessinger J., Downward J.,  
 Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";  
 RL Nature 309:418-425(1984).  
 RN 12  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN 13  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA MEDLINE=97078686; PubMed=8918811;  
 RA Reiter J.L., Maible N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN 14  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN 15  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX TISSUE=Placenta;  
 RA MEDLINE=21100872; PubMed=11161793;  
 RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Schehl C.M., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balaubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Maible N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Genomics 71:1-20(2001).  
 RN 16  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Thredgill D.W., Danielsen A.J., Schehl C.M.,  
 RA Lampland A.L., Balaubramaniam S., Crossley T.O., Magnuson T.R.,

RA Maible N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN 17  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells.";  
 RL Science 224:843-848(1984).  
 RN 18  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810(1984).  
 RN 19  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN 110  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";  
 RL Oncogene Res. 1:375-396(1987).  
 RN 111  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis.";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN 112  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN 113  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN 114  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 RT and supercoiled DNA.";  
 RL Nature 309:270-273(1984).  
 RN 115  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,  
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor.";  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN 116

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RA MEDLINE=96398132; PubMed=8962717;  
 RX Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts";  
 RL Growth Factors 13:121-132(1996).  
 RN (177)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668;  
 RA Sato C., Kim J.-H., Abe Y., Saico K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor";  
 RL J. Biochem. 127:65-72(2000).  
 RN (181)  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=9825196; PubMed=9556602;  
 RA Abe Y., Oda M., Inagaki F., Lax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor";  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN (191)  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens";  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X00588; CA25240.1; -  
 DR EMBL: U95089; AAB53063.1; -  
 DR EMBL: U48722; AAC50802.1; -  
 DR EMBL: U48723; AAC50804.1; -  
 DR EMBL: U48724; AAC50796.1; -  
 DR EMBL: U48725; AAC50797.1; -  
 DR EMBL: U48726; AAC50798.1; -  
 Query Match 46.4%; Score 3158; DB 1; Length 1210;  
 Best Local Similarity 49.8%; Pred.No. 7.8e-161;  
 Matches 631; Conservative 176; Mismatches 349; Indels 112; Gaps 23;  
 QY 11 LTLALPPGAA--STGVCTGDMKRLRPASPTHLDLRLYOGCCOVVQGNLELTYLPTN 68  
 DB 14 LLAALCPASRLAEKKVCQSTNKLTQVLGFEDHPLSLQMFNCEVVLGNLEITTVORN 73

QY 69 ASLSFLDIOEVOGYVLIANOVROVBLQRLIRVGTQLFEDNYALVLDNGDFLANNTP 128  
 DB 74 YDLSFKTIQEVAGVYLIANTVERIPILENQIIRGNMYENSALVLSNYD----- 126  
 QY 129 VTGASPEGLELQRLSTELTKGVLIQRNPOLCYQDTILMKQOYIANSFQITEL-- 186  
 DB 127 ---ANKTKLELPRNLOEILHGAVRSPNNPALCNVSIQWRD---IVSSDFLSNMGMD 180  
 QY 187 -NRSRACHPCSPMCKSGRCWGESEDCQSLTRVCAGGCA-RCKGPLPTDCCHQCAAG 244  
 DB 181 QNHGSCQKDPSCPCPNCSWGAGEBNCQTKITICAOQCGRCRCKSPSDCHQCAAGC 240  
 QY 245 TGRPHSDCLALFHNSGICELCPALVYNTDTPFESMPNDEGRYTFGASCVTCAPNYL 304  
 DB 241 TGPRESDCIVCRKFRDQATCKDTCPPMLNPTTYQMDVNDVEGKYSFGATCVKCKPRNV 300  
 QY 305 STDVSGTLVCPPLNOCVTVTAEDGTQREKSKPPARVCYGLAMHLEVRVATSNIOEF 364  
 DB 301 VTDHGS CVRACGADSYEM-EDGVKCKCKEGPCRVKCNIGIGEFDSLSINATNIXHF 359  
 QY 365 AGCKITFGSLAPLPESFPDGPASNTAPLOEQLOVFTEITGYLISAMPDLPDLS 424  
 DB 360 KNTCTISGDLHITPVAFRGDSFTHTPPLDQELDILKTVKELITGFLIQAHPNRTDLHA 419  
 QY 425 FQNLQVIRGRILHNQAVSLTLOGLISWLGRLSRLSGSLALIHNTHLCFVHTVPWQ 484  
 DB 420 FENLEIRGRTKQHGQSLAVSLNITSLRLSKETISDDDIVISGNKLCYANTIMWK 479  
 QY 485 LFRNPQALHTANRPDECVSGBLACHQLCANHCCKGPGPTQCVNCSQFLRQCYBEC 544  
 DB 480 LFGSGQKTIIRNNGENSCAKAGVCHALCSPEGCGPBPDCVSRNVRGREGVDKC 539  
 QY 545 RVLQGLPREVYNNRHLCPHPECOPOWNSYTCGPRADQCVACAHYDPPCVARCSGV 604  
 DB 540 NLGEGREPREVENSECIQCHPECLPQAMNITCTGRPDNCTIQCHAYIDGPPCVATPCAGV 599  
 QY 605 KPDLSTYPIWKFPDEBACQPCPINCTHSCVDDDKCQPAEORAPLSTISAVVG---1 661  
 DB 600 MGENNTL-VKRYADAGVCHLCHPNCTYCGTGGLBECPLTNGRIP--STATGVNALL 656  
 QY 662 LTVVLTGVFVGIILIKRQOKIRKRYTMRLLOETELVEPLTPSGAMPQAOQRILKETLR 721  
 DB 657 LTVVALGIG--LFMRRIHIVKRTLRRLQEBELVEPLTPSGEAPQALRIKETEFK 713  
 QY 722 KVVVLGSGAGTYKGIWIPDGENVKIPVIXYLRENTSKANETLDEAYVYAGVCSPP 781  
 DB 714 KIVVLGSGAGTYKGIWIPDGENVKIPVIXYLRENTSKANETLDEAYVYAGVCSPP 773  
 QY 782 VSRLLGICLTSTVQLVTLMPYGCILDHVENRGRGLSQDOLNMCMQIAGMSYLEDVRL 841  
 DB 774 VCRLLGICLTSTVQLITQLMPFGCLDYVREHNDIGSYLLNMKVQIAGMNTLEDRL 833  
 QY 842 VHRDLAARNLVKSPNHVKITDFTGLARLLDIDETEVYADGKYPKIMALESILRRFT 901  
 DB 834 VHRDLAARNLVKTPQGVKITDFTGLAKTLGAEEKEVYHAGKVPKIMALESILHRYTH 893  
 QY 902 QSDVWSYGVTVWELMTFGAKPYDGIIPARELPDLLEKGERLPORPICITDYMMIVKCM 961  
 DB 894 QSDVWSYGVTVWELMTFGSKRPYDGIIPASEISSLLEKGERLPORPICITDYMMIVKCM 953  
 QY 962 DSECRPRELVESEFARMARDPORFVYIQ-NEDLGAPSPDSTFFYSRLSDDDMGDLVDA 1020  
 DB 954 DADSRPKRRLIIEFSKWARDPORVIVIQGDEHMLPSPIDSNFYRLALMBEDMDVDVDA 1013  
 QY 1021 EBYLVPOQGFPCPDPAAGAGVHHRHRSSTSGGDLITLGLPSEEAAPSPPLASBEG 1080  
 DB 1014 DEYLIVPOQGF-----SSPSTRPPLSSLS 1039  
 QY 1081 AGSDVPDGLGMAAGLQSLPTHDPSPLORESDPTVPLPSET--DGYVAPLTCSPQPE 1138  
 DB 1040 ATSN--NSTVACIDRNGLOSCPIKEDSFLORYSSDPTGALTDSIDDTFL-----FVPE 1091



```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CONFLICT 19 19 C -> S (IN REF. 2)
FT CONFLICT 539 539 C -> W (IN REF. 5)
FT CONFLICT 991 991 L -> F (IN REF. 4)
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6)
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

```

```

Query Match 46.0%; Score 3136; DB 1; Length 1210;
Best Local Similarity 49.6%; Pred. No. 1,1e-159;
Matches 629; Conservative 171; Mismatches 366; Indels 102; Gaps 21;

```

```

QY 11 LLLALLPFGAA--STQVCTGTDMKLRPASBETHLDMRLHYQCCQVYQGNLELTYPTN 68
DB 14 LLLALLPFGAA--STQVCTGTDMKLRPASBETHLDMRLHYQCCQVYQGNLELTYPTN 73
QY 69 ASLSFLQDIQVQGVYLIANNQVQVPLQRLRIYRGTQLPEDNVALVLDNGDLNNTTP 128
DB 74 YDSFLKTIQVQGVYLIANNQVQVPLQRLRIYRGTQLPEDNVALVLDNGDLNNTTP 124
QY 129 VTGASPGRELEQLRSLEILKGGVLIQRNPOLCYDTIIMKQYIKANSKFIGITELNR 188
DB 125 -YGTNRTRGRLQELRLMRLQELIGAVRSPNPLCNMDTIQWRDIQVQVFMSPNMSMDQSH 183
QY 189 SRACHRECPMKSRCKSGESSEDCQSLTRVYVAGGCA-RCKGRLPTDCHEQCAAGTGP 247
DB 184 PSSCPKCDPSCPGSCWGGEGENCQKLTIKICAOQCCHRCGRSPSCCHQCAAGCTGP 243
QY 248 KHSDDLACLFHNSGICELCPALVTNTDTFESMPREGRYTPGASCTVACPNTYSTD 307
DB 244 RESDCLVQCFQDEBATCKOTCPRLMLNPTTYQMDVNPBGKISGATCVKCKCPNTYVTD 303
QY 308 VGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVYVAGMEHLREVAVTSANTIQEPAC 367
DB 304 HGSQVACGPDYEV--EEDGIRKCKCDGPCRKVCNGIGIGEPFDLTISINATNIKHKYC 362
QY 368 KRTFGSLAPLPESFPDPAANTAPLOEQVQFETLEITGYLYISAMPDLSPLDVSFON 427
DB 363 TALSGLHLPLVAFKDSFTRTPDLPRELEILKTVKEITGFLDIQAMPDNTDLHAFFN 422
QY 428 LOYIRGRILHNGVSLTLQGLGISWLGSLRLSLEGLALIHNNTHLCFHTVPMWDLFR 487
DB 423 LEIIRGRITKOHGFSLAVNGVLTSLGSLKLSIDSDVLTISGRNLCTVANTTNMKKLF 482
QY 488 NPHQALLHTANREDEDCVSEGLACHOLCARGHMGCPPTQCVNCSQPLRGQCEVEECRL 547
DB 483 TPQTKTKIMNRAEKQCKAVNHCNPLCSGSGMGPRPCVCSQVNSRRECEVECKNIL 542
QY 548 QGIPREYVNAARHLCPHRECPQWNGSVTCGPEADQCVACAHYKDPFCVACRPSGVKPD 607
DB 543 EGPREFEVENSEICQHPECLPQAMNITCTGRSPDNCIQCAHYIDGPHCVKTCPAGIMGE 602
QY 608 LSTYPMKFPDEGACOPCPINCTHSCVDLDDGCPAEGASPLSTISVAVNGIILVVL 667
DB 603 NNTL-VMKRADANNVCHLANCTYGCAGBGLQCEVWSPGPKIPSTANTIGVGLLFTIV 661
QY 668 GAVFGL-LIKRQOKIRKXTMRRLQETLEVBLPTSGAMPNOQMKILKTELARKVYL 726
DB 662 -VALGIGLPMRRHRIYKRTLRLLQERLEVEPLTSGEAPNQAHLRIKLETFKKIKYL 720
QY 727 GSGAFQTVYKGIWIPGENVKIVALKVLRNTSPKANIELDEAYVMAGVSPYVSRL 786

```

```

DB 721 GSGAFQTVYKGIWIPGENVKIVALKVLRNTSPKANIELDEAYVMAGVSPYVSRL 780
QY 787 GICLSTVQVLTQMPGICLLDHYENRGRGSGDLMCMQIKKMSYLEDVRLVHNDL 846
DB 781 GICLSTVQVLTQMPGICLLDHYENRGRGSGDLMCMQIKKMSYLEDVRLVHNDL 840
QY 847 AARNVLSPNHVIKTFGLARLIDIDETEHAGGKVPIMKMALESLIRRRFTQSDVM 906
DB 841 AARNVLSKTPQHVKITPGLAKLGAEEKEVHAGGKVPIMKMALESLIRRRFTQSDVM 900
QY 907 SYGTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVCMIDSECR 966
DB 901 SYGTWELMTFGSKPYDGIIPASDISILKEXGERLPQPICTIDVYIMVCMIDADSR 960
QY 967 PREPELSESRMARDQRFVTLQ-NEDLGASPLDSFFRSLLEDDEDMGLVDAEELV 1025
DB 961 PREPELSESRMARDQRFVTLQ-NEDLGASPLDSFFRSLLEDDEDMGLVDAEELV 1020
QY 1026 POGGFCPPDPAPGAGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEGAGSDV 1085
DB 1021 POGGFCPPDPAPGAGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEGAGSDV 1045
QY 1086 FDGDLGMAAGLQSLPTHPSPLORYSEDPVLPSET--DGVAPLTCSPOPEYVNP 1143
DB 1046 --NSTVACINRNSCRVKEDAFLORYSSDPTGAVTDNIDDAFL-----PVPEYVNP 1095
QY 1144 DVRRQPSRPGRLPAPRACATLERAKTISPGKGVVQVAFAGAVENPEYL-TPOGG 1202
DB 1096 SVRRPAGSVQNPVYHNPRLP-----APGDLHYON--PHSNVAGNPEYVNTAQ-- 1143
QY 1203 AAPQHPPPAPSPAFNDLYWMDQ-----DP-----PERGAPSTKGTPTENP 1246
DB 1144 -----PTLSSGFNSPALMIQKSHQMSLNDPVDYQDFFPKTKVNGIFKG-PTAENA 1195
QY 1247 EYLGDLVP 1254
DB 1196 EYLRVAPP 1203

```

RESULT 6

ERB4 HUMAN STANDARD; PRT; 1308 AA.

ID ERB4\_HUMAN

AC Q15303;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)

DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).

GN ERB4 OR HER4.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A. (ISOFORM JM-A).

RC TISSUE=Breast carcinoma;

RX MEDLINE=93189574; PubMed=838326;

RA Ploman G.D., Cullom J.-W., Whitney G.S., Green J.M., Carlton G.W.,

Foy L., Neubauer M.G., Shoyab M.,

RT "ligand-specific activation of HER4/p180erbB4, a fourth member of the

RT epidermal growth factor receptor family,"

RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).

RC TISSUE=Fetal brain;

RX MEDLINE=97476287; PubMed=9334263;

RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Ploman G.D.,

RA Klagesbrun M.,

RT "A novel junctional domain isoform of HER4/erbB4. Isoform-specific

RT tissue distribution and differential processing in response to

RT phorbol ester,"

RL J. Biol. Chem. 272:26761-26768(1997).

CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-  
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER  
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED  
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND  
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN  
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,  
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,  
 CC PLUTARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,  
 CC LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L07868; AAB59446.1; -;  
 DR HSSP; P11362; IFGK.  
 DR Genew; HGNC:3432; ERBB4.  
 DR MIM: 600543; -;  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 4.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 DR Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1308  
 FT DOMAIN 26 651  
 FT TRANSMEM 652 675  
 FT DOMAIN 676 1308  
 FT DOMAIN 186 334  
 FT DOMAIN 496 633  
 FT DOMAIN 718 985  
 FT NP\_BIND 724 732  
 FT BINDING 751 751  
 FT ACT\_SITE 843 843  
 FT DISULFID 189 197  
 FT DISULFID 193 205  
 FT DISULFID 213 221  
 FT DISULFID 217 229  
 FT DISULFID 230 238  
 FT DISULFID 234 246  
 FT DISULFID 249 258  
 FT DISULFID 262 289  
 FT DISULFID 293 304

FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 626 648 NGPSHDCIYYPWTKSHTLPGHA --> IGSSIEDICIGLMD  
 (IN ISOFORM JM-B).  
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;  
 Query Match 43.7%; Score 2975.5; DB 1; Length 1308;  
 Best Local Similarity 45.2%; Pred. No. 4,4e-151;  
 Matches 609; Conservative 185; Mismatches 379; Indels 173; Gaps 28;  
 9 WGLLLALLPGAA----STGYCTGTDMKLRLPASPEHLMDLRLYYGCGOVVGNLTLEY 64  
 8 WWWWVSLVLAAGTVQPSDSQSVCACTENKLSLSDLEQQYRALRRKYENCEVVMGMLTITS 67  
 65 LPTNASSLFLDIDIEVOCGYVLIANOVROVLOLRITRGQLFEDNVALAVLNDGDDL 124  
 68 IEHNRDLSFLRSVREYGVYVALNORFYLENLRIIRGTLYEDRYALAIPLVYRKDG 127  
 125 NTPVTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYOPTILMKDOYIANSKFGIT 184  
 128 NF-----GLDELGLAKNTLEILNGVYVDQKFLCYADTIHMODIVRPMPSNLTLY 178  
 185 ELNRSRACHPCSPMCKSRGCSSESDCQSLRTVACAGC-ARCKGPLPTDCHEQCAAG 243  
 179 STNSSGCGRCHKSCCTG-RCWGPTENHCQTLLRTVACABGCGRCVGPVSDCHREGCAG 237  
 244 CTGRKHSQCLALHPNNSGICELCPALVTVNTDTFESMPRPEGRTYTGASCYVACPVNY 303  
 238 CSGPRDTCFACMFMNSGACVTCQPTQFVFNPTTFLEHFMNAKYTGACVCKKCPNPF 297  
 304 LSTVGSCTVLCPLAHNOEVTADGTORCEKSKPCARCYGIGMEHLREVAVTSANIOE 363  
 298 V-VSSSCVTRACPSKMEV-ENENIKMKRCPCTDLCPRACDGIIGSLMSAQTVDSSNIDK 355  
 364 FAGCKKIFGSLAFIPESFDDPASNTAPLQPEOLQVETLEITGYLYISAMPDPLDS 423  
 356 FINCTKINGNLIPLVGTGHDPPYNAIEAIDEXKLVFRTVAIEIGFLNIQGWPMPTOFS 415  
 424 VFQNLQVIRGILNGAVSLTLOGLGSMGLRLSRLREGSLALIHNTHLCFVHTVWD 483  
 416 VFSNLVTIGGVLSGLSLTLTKOOGITSLQFOSLKEISAQNIYITDSSNLCYHTTWT 475  
 484 QLFNPHQALHTNARPEDECEGELACHOLCAGHCGMPPTQCVNCSQPLRGOCVEE 543  
 476 TLFSTINQRIYIRNRAENCTAGAWCNHLCSSDCGMRPDDCISRRRSKRICIES 535  
 544 CRVIGSLPREYVNAHCLPCHPECCP-ONGSVTCFGEPADQCVACAHYKDPFCVAPCPS 602

```

Db      536 CNLYDGEFREFENSGICVECDPOCEKMEBDGLTGHGPOPNCTCKSHKQPNVEKCPD 595
Qy      603 GVKPDLSTYMPYKFPDEEGACQPPINCTHSCVLDKGC-----PAEQRASPL 651
Db      596 GLQGANF--IFKYADPDRECHPCHPNCOTQCNPTSHDCIYFWPTGHSTLPQAR--TFL 652
Qy      652 TSIVSAV--GILVVVVGAVGGLIKRRQOKIRKRYTRRLLOEELVEPLTPSGAMPQA 710
Db      653 --IAAGYIGGLFIVLIVGLTFVAVYRRKSKIK-KRALRRFL--ETELVEPLTPSGTAPQA 708
Qy      711 QMRILKETELRKVAVLGSAGFVYKGMIPDGENKIPVAKYLRNTSPKAKEILDE 770
Db      709 QLRILKETELRKVAVLGSAGFVYKGMIPDGENKIPVAKYLRNTSPKAKEILDE 768
Qy      771 AYVAVGSPVYSRLGICLTSTVQVLTQMPYSCLLDHYENRGLSGODLIMWCQIA 830
Db      769 ALIMASMDHPLVRLVCLVCLSPITQVLTQMPHSCCLLEVYEHKONIGSQLLMWCQIA 828
Qy      831 KGMVSYLEDVRLVHRDLAARVAVLVSPPNHVKITDGLARLLDIDETEVHADGKPIKMA 890
Db      829 KGMVSYLEDVRLVHRDLAARVAVLVSPPNHVKITDGLARLLDIDETEVHADGKPIKMA 888
Qy      891 LESILRRFTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPICTID 950
Db      889 LECIHVRKFTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPICTID 948
Qy      951 VYVIMVYKCMVIDECPREFELVSEFSRMADPQRFVIONED-LGPASPLDSTFYKSL 1009
Db      949 VYVIMVYKCMVIDECPREFELVSEFSRMADPQRFVIONED-LGPASPLDSTFYKSL 1008
Qy      1010 EDDMDGLVDAEELVYPOQGFCDPAPAGAGVHHHRRSSSTRSGGDLTLGLEBSEE 1069
Db      1009 DEEDLEDMDAEELVY-QAFFNIPPP-----YTSHARIDSNMS-----EIGHSPPAY 1056
Qy      1070 APPS-----PLAP-SEGAGSDVPFDGLMGAKGLQS 1100
Db      1057 TPMSGNFVVRDGGFAAGGVSVYRAPSTIPEAPYAGATTAEIFDSCNGTLKRPVA 1116
Qy      1101 LPHDSEPLORYSEDPVLPIS-----ETDGVAVLTCSPQREYVNOQDVRPQSPSR 1153
Db      1117 PHVOESSTQRYSDADPTVFAFERSPRGELDEEGYMTMRKPKQEYLVNPE----- 1167
Qy      1154 EGPLPARPAGATLERAKTISPKNQGVKQVFAFGAVENPEYITPPCGAARPOHPPPA- 1212
Db      1168 ENDFVSR-----KNGDLO-----ALDPEYHNSNG-----PKAE 1199
Qy      1213 -----FSPAFDNLVYWDQDPPERGA--PPSTF 1237
Db      1200 DEVVNEPLVYNTFANTLTKAEYLKNNILSMPEKAKKAFDNPDIWNHSLPFRSTLQHPYL 1259
Qy      1238 KGTPT-----AENPEYL 1249
Db      1260 QEYSTKYFYKQNGRIRPIVAENPEYL 1285

```

```

RESULT 7
ERR4 RAT
ID ERR4 RAT STANDARD; PRT; 1308 AA.
AC 062956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulin promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269 (1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704 (1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF041838; AAD08899.1; -.
DR EMBL; U52531; AAC51051.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyRKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERB4-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```



```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT NP_BIND 718 985 PROTEIN KINASE.
FT BINDING 724 732 ATP (BY SIMILARITY).
FT ACT_SITE 751 751 ATP (BY SIMILARITY).
FT DISULFID 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB096A08B41 CRC64;
```

Query Match 43.4%; Score 2957; DB 1; Length 1308;  
Best Local Similarity 45.1%; Pred. No. 4.2e-150;  
Matches 608; Conservative 191; Mismatches 388; Indels 160; Gaps 28;

```
Qy 1 MEIA-ALCRWGLL--ALLPPGAATVCTGTDMKRLPASPETHLDMRLHLYOGCQVQ 57
Db 1 MKLATGLMWGSLVLAARTVQPSASQSVCACTENKLSLSDLEQYALARKYENCENV 60
Qy 58 GNLELYLPTNASLFIADIOEVGYLIAHNOVRQVPLQRLVIRGTQLEFEDYVALVL 117
Db 61 GNEIETSIENHRDLSFLRSIREVGYVALNQFRYPLLENLRIRIRGKLYEDYALAI 120
Qy 118 DNGDPLNNTPTVYGASDGLRELOLRSLTEILKGVLIQRNPQCYQDITIMKOYIKAN 177
Db 121 LNRKQDNF-----GLOELGKNTLTELINGGVYVDQNKFCYADTIHMVDI VNPW 171
Qy 178 SKFGLTELNRSPRACHPCSPWCKSGRCGSESSDQSLTRTVCAAGC-ARCKGLPTDCC 236
Db 172 PSNMTIVSTIGSSCGRCHKSCG-RCMGPTENHCQTLTRTVCAEQDGRGYPVSDCC 230
Qy 237 HEQCACTGPKSHDCLACHFNHSGICELHCPALVYNTDTFESMNPBGRYTFGASC 296
Db 231 HRECAGGSGPKDIDCFACNMNFDSGACVYQCPOTFYNNPTFLEHNFNAKYTYGAF 290
Qy 297 TACPYNLTSDVSGCTLVCPLNQEVTAEDGTORCEKSKRCARVYGLGMEHLREVR 356
Db 291 KKCPHNIV-VDSSSCVAAACBSKMEV-ENNGIKCKECTDLCIPACDGIIGSLMSAQ 348
```

```
Qy 357 TSANIOEACCKKIFGSLAFLPESFDGDPASNTAPLPEOLOVETLEETIGYVISAP 416
Db 349 DSSNIDRFNCTKINGLIFLVYGHDPDPAIDAIDPEKLVNRYTRREITGFLNIQWP 408
Qy 417 DSLPDLVFNOLVIRGRILHNGAYSLTLOGISWLSGLSLRELGLALIHNNHLCF 476
Db 409 PNMTDFSVFNVLVIGRVLVYSGSLILLKQGGITSLQFQSLKEISAGNIYITDMSNLC 468
Qy 477 VHTVPMQQLRRNPQALHTANRPEDCEVGBGLACHOLCARGHGWGPGPOCVNCSQFL 536
Db 469 YHTNMWTLFSTVYQRLVIRIDNRRAENCTAEAGVCMNLCSDNGCWGPGPDCLSCRRFSR 528
Qy 537 GOECVEECRVLOGLPREYVNAHCLPCHPECP-ONGSVTCFGEADQCVACAYKDPF 595
Db 529 GKICIESCNLYDEFFREFNGSICVEEDSQCEKMEDELITHGCPDPNCTCSHFQCPN 588
Qy 596 CVARCPGVKPDLSYMBIMKFPDEGACQCPINCTHSVDLDDKC-----PA 644
Db 589 CVEKCPVLQGANF--IFKADODRCHPCHPCTGCGNGPSTSHDCIYPMTGHSLLPQ 646
Qy 645 EGRASPLTYSANV-GILVVVLGVVFGILIKRQOKIKRYTKRRLLOETELVEPLTPS 703
Db 647 HAR-TPL--IAAGVIGLFLVIMALTFAVYVRRKSIK-KRRALRRFL-ETELVEPLTPS 701
Qy 704 GAMPNQAOMRLIKETELRKVYLGSGAFGYKGIWPDGENVKI PVAIKVLRNTSPKA 763
Db 702 GTAPNQAOLIKETELRKVYLGSGAFGYKGIWPDGETVVI PAIAIKLNETTGPKA 761
Qy 764 NKEILDEAYVAVAGVSPYVRLLGICLTSTVOLVTQIMPYGLLDHVENRGRIGSODL 823
Db 762 NVEEMEDALIMASVDHDLVRLGVCSPITQVLTQMPHCILLEYVHEHKONIGSOLL 821
Qy 824 NMCMQIAKMSYLEADVLYHRDLAARVLYKSPHNVKITDQGLARLIDISTEYHAQCGK 883
Db 822 NMCMQIAKMSYLEADVLYHRDLAARVLYKSPHNVKITDQGLARLIDISTEYHAQCGK 881
Qy 884 VPIKMALESILRRFTHOSDVMYGYVWELMFFGAKPYDGI PAREIPDLLEGERLPQ 943
Db 882 MPIMKMALECIHRKFTHOSDVMYGYVWELMFFGAKPYDGI TRREIPDLLEGERLPQ 941
Qy 944 PPICTIDVYIMVCMWIDSECRPRFRELVSFESRMARDPQRFVYIONED-LGPASPLDS 1002
Db 942 PPICTIDVYIMVCMWIDSECRPRFRELVSFESRMARDPQRFVYIONED-LGPASPLDS 1001
Qy 1003 TFRSLLEDDEDMGULVABEYLVPOQGFPCDP-----APGA 1039
Db 1002 KFFQNLLEDDEDMGULVABEYLVPOQGFPCDP-----APGA 1038
Qy 1040 GGMVHHRHRSSTFRSGGDLTLGLEPSEEARPRSLPASREGAGSDVFDGLGMGAAGLQ 1099
Db 1061 GSGFVYDGGFATQGG---MPMTYATTSITLIPAPVA--QCATREMFDSDSCNGTLRKPV 1115
Qy 1100 SLPTHDSPLQRYSEDPVLPSS-----ETDGYVALPCLSPQREYVNOVDVAPQPS 1152
Db 1116 VPHVQEDSSTQRYSADEPTVAPERNPRAIDEBEGYVPMHDKPKQETLNPVE----- 1167
Qy 1153 REGPLPARAGATLEAKTLSPGKNGVADVAFGGAVENTPEYLTLTQGGAAAPPHPPA 1212
Db 1168 -ENPFVSR-----KNGDLQ-----ALDNPEVHSSASG-----PPKA 1198
Qy 1213 -----FSPAFNLVYWDQDPERGA--PPT 1236
Db 1199 EDEVNEPPLVNLFTNALGNAEYMKNSLSVPEKAKAFAFNDPDMHNSLPRRSTLQHPDY 1258
Qy 1237 FKGTPT-----AENPEYL 1249
Db 1259 LOEYSTKYFYKQNGRIRPIVAENPEYL 1285
RESULT 8
XMRK_X1PMA STANDARD; PRT; 1167 AA.
ID_XMRK_X1PMA
AC_P13388;
```



```

DT 01-JAN-990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TY.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=90015140; Pubmed=2797166;
RA Wiltbrodt J., Adam D., Malitschek B., Mauelel W., Raulf F.,
RT Telling A., Robertson S.M., Scharl M.;
RA "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing ty locus in xiphophorus.",
RN Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Scharl M.;
RL Submitted (JUL-2000) to the EMBL/Genbank/DDBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X16891; CAA34770.2; -.
DR PIR, S06142; S06142.
DR HSP, P11362; 1R6K.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR Prodom; PD000001; Euk_pkinase; 1.
DR SMART; SMO0261; FU_5.
DR SMART; SMO0220; S_TKc; 1.
DR SMART; SMO0219; TYKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167
FT DOMAIN 26 642
FT TRANSEM 643 665
FT DOMAIN 666 1167
FT DOMAIN 710 977
FT NP_BIND 716 724
FT BINDING 743 743
FT ACT_SITE 835 835
FT DISULFID 195 204
FT DISULFID 199 212
FT DISULFID 220 228
FT DISULFID 224 236

```

Query Match	Beet Local Similarity	Score	DB 1	Length	1167
Matches 574; Conservative 165; Mismatches 391; Indels 137; Gaps 27;					
FT DISULFID	237	245	BY SIMILARITY.		
FT DISULFID	251	253	BY SIMILARITY.		
FT DISULFID	256	265	BY SIMILARITY.		
FT DISULFID	269	296	BY SIMILARITY.		
FT DISULFID	300	311	BY SIMILARITY.		
FT DISULFID	315	330	BY SIMILARITY.		
FT DISULFID	333	337	BY SIMILARITY.		
FT DISULFID	504	513	BY SIMILARITY.		
FT DISULFID	508	521	BY SIMILARITY.		
FT DISULFID	524	533	BY SIMILARITY.		
FT DISULFID	537	553	BY SIMILARITY.		
FT DISULFID	556	569	BY SIMILARITY.		
FT DISULFID	560	577	BY SIMILARITY.		
FT DISULFID	593	615	BY SIMILARITY.		
FT DISULFID	618	626	BY SIMILARITY.		
FT CARBOHYD	114	114	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	144	144	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	201	201	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	356	356	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	365	365	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	398	398	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	417	417	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	501	501	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	576	576	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
FT CARBOHYD	621	621	N-LINKED (GLCNAC . . .)	(POTENTIAL).	
SEQ	1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;				
Query Match	39.7%;	Score 2700.5;	DB 1;	Length 1167;	
Beet Local Similarity	45.3%;	Pred. No. 1.7e-116;			
Matches 574; Conservative 165; Mismatches 391; Indels 137; Gaps 27;					
4 AALCRMGILALALPEGAAS---	QVCTGTDMLRLPASPETHIDMLRHLYOGQOVQGN	59			
8 AALAQ--LLVLVLSIRCCSDPDPRKVKQOGTSMQTM--	LDNHLYLKKMKMYSGCNVLEN	62			
60 LELTYLPTNASLSFLDIOEVQGYVLIAHQVROVPLRLRIYRGTLFEDNYALAVLDN	119				
63 LEITYTQENDQSLFQISQEVGVLLIANNEVSRPIPLVNLRLIRGQVLYEGNFTLVMSN	122				
120 GDPILNTTPYTGASPGGLRELOLRSLTEILKGYVLIGRNPOLCYODTILMKDOYIKANSK	179				
123 YQK-NPSSP--DVYQVVLKQLOLSNLTLEILSGGVKVSHPNLLCNVETINMMDIVDKTSNP	179				
180 FIGITELNRSRACHPCSPMKSGSRCKSESESDOSLRLTYCAGGC-ARCGRLPTOCHE	238				
180 TMLIPIPAFERQCKCHGCVNGSCVAPRGHCKFKLLCAQCCNRRCRGPRIIDCNE	239				
239 QCAAGCTGPKHSDCLAFHNSGICGLCPALVTVTTDTFESMPNPEGRTYGASCVTA	298				
240 HCAGCGTGPATDCLACRDFNDSDGTCMDTTPPKIYDIYHQVYVNDPNIKYTTGAAVCKE	299				
299 CPVNYLSTDVSGCTLVLCPLHNOEVLVAEDGTQRCCKSKPCARVYCYGLGMEHLREVAVYS	358				
300 CPSSNYVYTE-GACVRSQCSAGMLEVD-ENGKRSCKPCDGVCPKVDGIGISLNTIAVNS	357				
359 ANIOEPAGCKKIGISLAFLPESFDGPASTAPLOLBOLQVETLEBTLEITYLIASMPDS	418				
358 TNRISFENCKIKINGDIIILNRSFEGGPHYKIGTMDPEHLMNLITVVEITYGLVIMMPEN	417				
419 LPDLVSQNLQVIRGRILIHNGAYS-LTLQSLGISMILGRILRELGSLALIHNNTHLCFV	477				
418 MTLISVQNLIEIRGRITTPSRGFSFVVVQVYRHLQMLGSLNSKVSAGNVLTKNTLOIRYA	477				
478 HTPVMDOLFENPHQALLHTANRDEDCVEGGLACHOLCARGHCMGPGPTQCVNCSOFLRG	537				
478 NTINMRILFRFESDOSIETDART-----ENQCNNECSDEGCMGPGPTCVSCLAHVDG	530				
538 QECVEBEVRVLQGLPREVYVNAHCLPCHPBGQOPQNSVYTCGPRAADQCVACAHYKDPFCV	597				
531 GRCAVACSNLQGGPRRAQVNGRVCVOCHQDELTVQTDLTCTGPGPANCSSAHFQDDPQCI	590				
598 ARCPGSGKPLSLVMPILMKPFDEGACQPCPQINQTHSCYULDKDGCAFAFARASPLSTIVA	657				

```

Db      591 PRCPHGILGSDGDTL-IMKYADMGCGCQCHQNCCTGCGSGPGLSGCRGD-IVSHSLAVGL 648
Qy      658 VVGLLVVGVGVGILIKRQOKRKRTMRLLDQETLVEPLTSGAMPNQAQRIIKE 717
Db      649 VSGLLITYIVALLVYLLRRRRIRK-RKRTIRCLLOEKELVEPLTSGCAPNQAFLRIIKE 707
Qy      718 TELRKVKVLGSGAGFCTVYKGIWIPDGENVKI PVAIKVIRENTPANKNEIIDEAVVMAGV 777
Db      708 TEFKDKRVLGSGAGCTVYKGLMNPGENIRIPVAIKVIREATSPYKNOVEIDEAVVMASV 767
Qy      778 GSPVYSRLLAGICTSTVOLVTLMPYGLLDVHRENRGRLSQDLLNMCQIAKMSYLE 837
Db      768 DHPVYCRLLGICTSTAVQTLVTLMPYGLLDVYRQHQRICQOMLLNMCVQIAKMANYLE 827
Qy      838 DVRLVRLAARNVYKSPNNVKTIDPGLARLDDIDETEHADGKGVYIKMALESIIAR 897
Db      828 ERHLVHRDLAARNVYLLKNPNVKTIDFGLSKLLTDADEKEYOADGKGVYIKMALESIIQW 887
Qy      898 RFTHOSDVMSYGVTVWEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWK 957
Db      888 TYTHOSDVMSYGVTVWEIMTFGSKPYDGIIPAKEIASVLENGERLPQPICTIEVMITLK 947
Qy      958 CWMIDSECRPRPRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILDEDDMDL 1017
Db      948 CWMIDPSRPRFRELVEFGSQMARDPKRYLVIG--NLPSLSDRLFSRLSSD--DV 1002
Qy      1018 VDAEYVLPQGFPCFPDPAAGCMVHRHSSSTRSGGDLTLGLFSESEAPSPAP 1077
Db      1003 VDAEYVLPYKRI-----NRQGS-----EFCIF 1025
Qy      1078 SEGAGSDVFDGDLGCAKGLQSLPTHDPSPLORYSEPTV-PLPSETDGVYAPLTSPQ 1136
Db      1026 PTGH-----PVRENSITLRIISPTQNALEKDLDG----- 1056
Qy      1137 PEYVNOQPVPRP-----DSPRE-----GPLP-AABPAGATLERAKTISPKXGVYKD 1183
Db      1057 -EYVNOQFSETSSRLSDIYNPYEDLTGQWGVSSISOEAETNFSRPEVLTNQNLSL--- 1112
Qy      1184 VFARGAVENPEVYLPQGAAPQHPHPAPFAPFNLYVWDDPDEGAPSTFPGCTPTA 1243
Db      1113 PLVSSGSHDDDPY---QNG-----YQNAF-----LPQTALTGNGMFLPAA 1150
Qy      1244 ENPEYLG 1250
Db      1151 ENLEYLG 1157

RESULT 9
ERB3_HUMAN STANDARD; PRT; 1342 AA.
ID ERB3_HUMAN
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Pomescu N.C., Aaronson S.A.;
RT Isolation and characterization of ERB3, a third member of the
RT ERB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

```

```

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE ERB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS. A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AA35790.1; -;
DR EMBL; M34309; AA35979.1; -;
DR EMBL; S61953; AAB26935.1; -;
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Gene; HGNC:3431; ERB3.
DR MIM; 190151; -;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SMO0261; FU; 3.
DR SMART; SMO0219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP. FALSE NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR. FALSE NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Trnsmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1..19
FT CHAIN 20..1342
FT DOMAIN 20..643
FT TRANSMEM 644..664
FT DOMAIN 665..1342
FT DOMAIN 709..966
FT NP_BIND 715..723
FT BINDING 742..742
FT ACT_SITE 834..834
FT DISULFD 186..194
FT DISULFD 190..202
FT POTENTIAL.
FT RECEPTOR PROTEIN-TYROSINE KINASE ERB3-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN KINASE.
FT ATP (BY SIMILARITY).
FT ATP (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.

```

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDPLCHMDTIDMDIYDRDAEIVVQNGR
SC -> GQFVWVSGLTLPQPDQWYLLDDDRLLTSSASK
VPVTLAAV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

```

Query Match 35.7%; Score 2432.5; DB 1; Length 1342;  
 Best Local Similarity 40.7%; Pred. No. 3.6e-122;  
 Matches 534; Conservative 191; Mismatches 457; Indels 129; Gaps 32;

```

QY 10 GILLALLPFGAA--STOVCTGTOMKLLRPASPEHLDMLRHLYOGCGVQGNLELTLYPT 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 GLLFSLARGSEVGNISQAVCPGTLNGLSVTGDAENQYOTLYKLYERCEVWGNLEIVLTGH 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 68 NASLSFLODIOEVQGYVLIHANQVQVPLRLRIVRGTOLFEEDYVALAVLNDGDPINNTT 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 NADLSFLQWIREVTGYVLVAMNERSTLPENLRVVRGTQYDGFALFVM-----LNYNT 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 PYTGASPGGLRELQLRSLEILKGVLIQGNPOLCYODTILMKDQYIKANSKPIGITELN 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDMDIYDRDAE---IVVKD 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 188 RSRACHPCSPMGCSRCWSSSEDCSLTRTVCAAGC-ARCKGRLPDDCCHBOCAAGCTG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 NGSSCPCEVCHGK-RCMWGSSEDCQTLTKTICAPQCNHGFQGNPNQCHDEACGCSG 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 PKASDCLACLHFNHSGICELHCPALVYNTDTFESMNPGRGYTFGASCVTACPYNYLST 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 PQQTDCAFCHRFYDSGACVRCQPLVYNTLTQLLENPHTKIYQGVGVCVASCAPHNV-V 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 DVGSCCTLVCPHNOEVTAEADGTORCEKSPCARVCYGLGMEHLREVRAYTSANTIOFAG 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 DQTSCTACACPPDKMEVD-KNGLKMCCEPCGLCPACSGTSG--SRQYDSSNIDGFVN 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 CKKIFGSLAFPSFGDPASNTAPLQPELOVETLEETITGYLYISAWPDSLPLSVFQ 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 CKILNDLDELITGLNGDPMHKTALDPEKLVNFRVYREITGYLNTQSWPMHMFVSF 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 NLOVINGRIHNGAVS-LTLOGLGISWLGRLSRELQSGGLALIHNNHTLCFVHTVPDOL 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 NLTTIGRSLYNGRFSILMKNLNVTSLGFRSLKEISAGRIYISANROLCTHHSLNMTKY 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 486 FRNPQALLHTA-NRPEDECVGEGLACHQQLCARGHCWGPPTQCVNCSQFLRGCEVSEC 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 LRGTETRLDICKNRPRRDCVAEGKVCDDLCSGSGCWPBGQCCLSCNYSRGVCVTHC 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 545 RVLOGLREYVNAHCLPCHPRECCPQNGSVYCFPREADQCVACAHYDPPRCVARCPSGV 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 NPLNGEPRERPAHEDECSCHPECQPMEGTATCNCSGSDTCAQCAHFRGPPCVSSCPGV 593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 605 KPDLSYMPIMKFPDEGACOPCPINCTHSVDLDDKCGCPABQRA----SPITSIVSAVVG 660
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 594 LG--AKRPITYKPRVQNECRCHENCHTQCGCPRELQDCLGTLVLIGTHTLMTLVLAG 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 ILVVVLGVVFGILLKRRQKIR-KYTMRLLOETELVEPLTPSGAMPNOAMRLKETE 719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 --LVVILFMILGTFLYRGRRIQKRAMRYLERGESIEPLDPS-EKANIKVLARIFKETE 708
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 720 LRKKVUGSGAFGVYNGIWMPEDENKIPALIVLEBNTSPKANKELIDEAIVVMAGVS 779
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 LRKLKVLGSGVFGVHKGWIPGESIKIPVCIVIEDKSGROSFOAVTDHMLAIGSLDH 768
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 780 PYVSRLLGICLTSTVQVLTOLMPYGCLLDHVRENRGRSGQDLNMCQIAKMSYLEDV 839
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AHVRLGLCPGSSLOLVTOYTLPLGSLDHVKQRGALGOLLNMGVQIAKGMVYLEEH 828
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 840 RLYVRDLAARNVLVKSBNHYKITDFGLARLIDIDETEHADGKVPYIKMALLESILRRRF 899
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 GMVHRNLAAARNVLKSPSQVADFVADLLPPQDKQLLYSEAKTPIKMALESIHFEKY 888
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 900 THQSDVMSYGVTTWELMTFGAKPYDGI PAEIRPLLEKGERLPORPCTIDVYIMVYCW 959
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 THQSDVMSYGVTTWELMTFGAEPYAGRLAEVPRLLKEKGERLAQPOICTIDVYIMVYCW 948
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 960 MIISECPRPRELSEFSRMAPDPORFVUIONEDLGA---SPDSTFYRSLDDDDMGD 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 MIDENIRPTELKELANETFRARDPRLVYIKRS-GEGLAPRPFGTLTKKLEVELEER 1007
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1017 LVDAEYLVPQCGFFCPDPAPGAGVWHHRSSSTSGGDLTLGLER-SEEAAPSPL 1075
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1008 ELDDLDLLEABED-----NATTTLSALSLPVGTNLRPGSGSL 1048
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1076 APSEGAGSDVFDGDLGKAAGLQSLPETHD-PSFLQRYSEDPVPLP-----SETDGYV 1128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1049 SPSGSG-MPMNOGLTGSCQESAVSGSSERCPRVSLH-----PMRGLCLASSESECHV 1101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1129 A-----PLCSPOPE-----YVNGPDVRPQPPSPRPP-----L 1157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1102 TGSEAEILOEVYMSCRSRSRSPRPGRDSAHNSQHSLLTPVTVLSPPGLEEDVNGYVM 1161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1158 PAARPAGATLERAKTISP-GKNGV-----KDFAFGCAVENPEYLTPOGGAAPORPAP 1210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 PDTHLAKETPSSREGTSSVGLSSVLTGTEDEED-----EEYEVNRRRRHSP-PHPP 1212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1211 PAFSPAFLNLYWD-----QDPERGAPSTFKGTPTAENPEYL 1249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1213 RPSLEELGYEYMDVGSLSLSIGSTQSCPLHPVPIWPTAGTTPDEDEYEV 1263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10  
 ERB3\_RAT  
 ID\_ERB3\_RAT STANDARD; PRT; 1339 AA.  
 AC 062799; 062955;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase etdb-3 precursor (EC 2.7.1.112)  
 GN (c-erbB3).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=96096535; PubMed=8522190;  
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.,  
RT "Cloning of the rat ErbB3 CDNA and characterization of the  
recombinant protein.";  
RL Gene 165:279-284 (1995).  
RN (2)  
RP REVISIONS TO 85: 513 AND 565.  
RA Hellyer N.J., Koland J.G.,  
RN Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.  
RL (3)  
RP SEQUENCE OF 922-1097 FROM N.A.  
RX STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
RC MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.,  
RT "Expression of neurogins and their putative receptors, ErbB2 and  
ErbB3, is induced during Wallerian degeneration.";  
RL J. Neurosci. 17:1642-1659 (1997).  
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAR.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
(POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC -1- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
PHOSPHATIDYLINOSITOL 3-KINASE.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U29339; AAC28498.2; -  
DR EMBL: U52530; AAC53050.1; -  
DR HSBP; P11362; IFGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU\_5.  
DR SMART; SM0219; TyKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP\_FALSE\_NEG.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 19  
FT CHAIN 20 1339  
FT DOMAIN 20 643  
FT TRANSSEM 644 662  
FT DOMAIN 663 1339  
FT DOMAIN 183 259  
FT DOMAIN 707 964  
FT NP\_BIND 713 721  
FT BINDING 740 740  
FT ACT\_SITE 832 832  
FT DISULFID 186 194  
FT DISULFID 190 202  
FT DISULFID 210 218  
FT DISULFID 214 226  
FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.  
FT DISULFID 246 255 BY SIMILARITY.  
FT DISULFID 259 286 BY SIMILARITY.  
FT DISULFID 290 301 BY SIMILARITY.  
FT DISULFID 305 320 BY SIMILARITY.  
FT DISULFID 323 327 BY SIMILARITY.  
FT DISULFID 500 509 BY SIMILARITY.  
FT DISULFID 504 517 BY SIMILARITY.  
FT DISULFID 520 529 BY SIMILARITY.  
FT DISULFID 533 549 BY SIMILARITY.  
FT DISULFID 556 573 BY SIMILARITY.  
FT DISULFID 576 585 BY SIMILARITY.  
FT DISULFID 589 610 BY SIMILARITY.  
FT DISULFID 613 621 BY SIMILARITY.  
FT DISULFID 617 629 BY SIMILARITY.  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 1028 1028 L -> P (IN REF. 3).  
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BFPD1E CRC64;  
Query Match 34.7%; Score 2362.5; DB 1; Length 1339;  
Best Local Similarity 41.0%; Pred. No. 1.9e-118;  
Matches 526; Conservative 172; Mismatches 430; Indels 155; Gaps 35;  
3 LAALCRNGLLALPPGAA---STOVCTGTOMKRLRPSRETHLMDLHLYOGCOVQGN 59  
7 LQVLC---FILSLARGSEMGNSQAVCPGLTNGSLVGTGDANQYOTLYKLYEKGVNMGN 62  
60 LELTYLPFNASLFLQDIOEVQGYVLAHNOVROPLQRLRIYVGTQLFEDNYALAVLDN 119  
63 LEIVLTGHNADLSFLQWIREVTGYVLAHNFVSFLPRLNLAHVNGTQYDDKFAIFV-- 120  
120 GDPPLANTPTVTGASPGGLREQLRLSLTEILKGVLIQRPOLCYDTITLMDQYIKANSK 179  
121 ---LNTYNT---NSSHALRQLKFTQLTEILSGVYIEKNDKLCMDITDMD-IVRVGA 172  
180 FIGITELNRSRACHPCSPCKSGKSCWKESSDCOSTRTTVAGGC-ARCKGPLPDDCHE 238  
173 EIVVK--NNGANCPKCEVCKG-RCKWGGPDDCQILTKTICAPQCNCRCPNPNQCCHD 229  
239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNREGRTYFGASCYTA 298  
230 ECAGGCGSPQDTDFACRRFNDSGACVRCPEPLVYNKLTQLEBNPHTKYQYGVCVAS 289  
299 CPVNYLTDVSGCTLVCPRLHNOEVTAEEDGTORCEKSKPCARVCYGLMEHLREYRAVTS 358  
290 CPNHFV-VDQFFCYRACPPDMEDVD-KHGLMKCEPCGGLCPKACEGTSG--SRQYVDS 345  
359 ANIEFAPGCKKIPGSLAFPPSPGDPASNTAPLQPELOVFETLEETGLVLYISAWDS 418  
346 SNIGFNVCKTILNLDLTLGLNVDPWHKIPALDPELVNFRVREITGLYNTQSWPPH 405  
419 LPDLVSFQNLQVIRGRILHNGAVS-LTLQGLGISWLGARSIRELSGLALIHNTHTLCEV 477  
406 MHNVSIVSNLTLTIGRSLYNNGFSLIMKNLNTVSLGRSLKETSAGAVVISANQOLCYH 465  
478 HTVPMDQLFENPHQALLHTA-NRPEDCVGEGALCHQICARGHCWGPSPGTCVNSQFLR 536  
466 HSLMTTRLLRPSSEERLDIKYDRPLGECLAGKVCDDPLCSSGSGCGPSPGCLCRNYSR 525  
537 GQECVSEGRVYQGLPREVYVARHCLPCHPRECQPNQSGVTCGSPADQCVAAHYKDPFC 596  
526 EGVCTVHCNPLQGGPREVYHQAQCPSCPECLPMEGSTCGSGSDACARCAHFRDGHG 585  
597 VARCPGKVPDLSTYMPITWKPPDEGACOPCPINCTHSC--VDLDDKCGPAPQASPLTIS 654



RT growth factor receptor.";  
RL Nature 314:178-180(1985).  
RN [8]  
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
RP ANALYSIS.  
RX MEDLINE=92038942; PubMed=1936959;  
RA Raz E., Schejter E.D., Shilo B.Z.;  
RT "interallelic complementation among DER/rlb alleles: implications for  
RL the mechanism of signal transduction by receptor-tyrosine kinases.";  
RN Genetics 129:191-201(1991).  
RP [9]  
RP REVIEW.  
RX MEDLINE=97248481; PubMed=9094709;  
RA Perrimon N., Perkins L.A.;  
RT "There must be 50 ways to rule the signal: the case of the Drosophila  
RT EGF receptor.";  
RL Cell 89:13-16(1997).  
CC -I- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARMO, WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-MARK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS. CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF ANTIOSEROSA AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF CUTICLE.  
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
CC -I- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE PROTEIN.  
CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -I- TISSUE SPECIFICITY: UNBOUTIOUSLY EXPRESSED IN EMBRYOS, IN LARVAE, UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLGEN OF TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX AND THORACIC AND ABDOMINAL GANGLIA.  
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF052754; AAC08536.1; -  
DR EMBL; AF052753; AAC08536.1; JOINED.  
DR EMBL; AF052754; AAC08536.1;  
DR EMBL; AF052753; AAC08535.1; JOINED.  
DR EMBL; K03054; AAAS1460.1; -  
DR EMBL; K03417; AAAS1460.1; -  
DR EMBL; K03416; AAAS0965.1; -  
DR EMBL; K03418; AAAS1461.1; -  
DR EMBL; AF109077; AAD26132.1; -  
DR EMBL; AF109078; AAD26132.1; -  
DR EMBL; AF109082; AAD26132.1; JOINED.  
DR EMBL; AF109078; AAD26133.1; -  
DR EMBL; AF109084; AAD26133.1; JOINED.  
DR EMBL; AF109079; AAD26130.1; -  
DR EMBL; AF109081; AAD26130.1; JOINED.  
DR EMBL; AF109079; AAD26131.1; -  
DR EMBL; AF109083; AAD26131.1; JOINED.  
DR EMBL; AF109080; AAD26135.1; -  
DR EMBL; AE003454; AAF46732.1; -  
DR EMBL; X02293; CAA26157.1; -  
DR EMBL; X78920; CAA55523.1; -  
DR EMBL; X78918; CAA55521.1; -

DR EMBL; X78919; CAA55522.1; -  
DR PIR; A00640; GQPE.  
DR HSSP; P13362; 1EGK.  
DR FlyBase; FBgn0003731; Egfr.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase.1.  
DR Pfam; PF00757; Furin-like.1.  
DR Pfam; PF01030; Recep\_L\_domain.2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 7.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase; Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing; Developmental protein.  
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing; Developmental protein.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.  
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 869 889 POTENTIAL.  
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 938 1198 PROTEIN KINASE.  
FT NP\_BIND 944 952 ATP (BY SIMILARITY).  
FT BINDING 971 971 ATP (BY SIMILARITY).  
FT ACT\_SITE 1063 1063 BY SIMILARITY.  
FT MOD\_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
FT ACT\_SITE 902 902  
Query Match 28.9%; Score 1967; DB 1; Length 1426;  
Best Local Similarity 32.8%; Pred. No. 2; 3e-97;  
Matches 467; Conservative 186; Mismatches 433; Indels 338; Gaps 41;  
24 QVCTGTDMKRLPASPETHLMLRLHYGCGVQGNLELYPT-NASLSFLQDIQEOVG 82  
100 KICGTGSRRLVSPSKHNNLRNYNCTVQGNLTLPLNPNLDSFLDNIREYTG 159  
83 YVLAHNOVROYPLORLRIYVGTOLF-----EDNALVALVNDGNPLNNTPYTGASPGGL 137  
160 YILSHVDVKKVPPKQIIRGLTFLSVSEKVALFV-----TYSKM 203  
138 RELQRLSTELKGGVLLQRRPOLCYDPTILMKQOYIANSKFIGITELNSRACHPSP 197  
204 YTLRPPDRDLVNGVGHNNYNLCHMRTIQSEIVSNGTDAYNYDTPARRECPKGE 263  
198 MCKGSRGWESSESDQSLTRTVACAGCA--RCKGPLPTDCGHEQCAAGCTGPKHSIDLAC 255  
264 SCTHG-CWGEGRPNQKFKSKYLCSQCAAGRCYGRPKRECHLFCAGGCTGTGTDCTIAC 322  
256 LHPNHSICELHCPALVTYNTDTPESMPNPGRYTFGASCYATACPYNYLSTDVSGCTIWC 315  
323 KNFDEAVSKKECPMRKYNPTTYLLETNPBGKYAVGATCYVECP--GHLLRDNGACVSC 381  
316 PLHNGEVTABDGTORCEKSCPKARVCYGLGMEHLREVRATYSANIOFAGCKIFGSLA 375  
382 PDKMKDKGE-----CVPDNGPCPTKCPGVTLH-----AGNIDSFNCTVVDGNIR 428  
376 FLPSFDS--DPASNTA-----PLOPEQLQVFTLEETITGLYLSAMPDLSPLDSVQN 427  
429 ILDTGFSGFQDVVANYTWGPRYIPLDPERREAVFSTVKITGLYLNIEGHHPGRNLSYRN 488  
428 LQVRIGRLIHNGAY-SLTLOGLIGISWGLRSLRBLGSLALIHNTHLCPHTVYPMDF 486  
489 LETIHGRQLMESMPAALAIYVSSYLEMRILKQISSGSVVIQHNRDLCYVSNIRFWPIQ 548  
487 RNPQALLHTANRPEDECVGEGGLCHQICARGHCGPGPTQCVNCSQPLRQGECEBGRV 546  
549 KEPQKQVAVNMLPADLCEKKGITICSDQCNEDCGAGACTDCLTCXKNPNFGTCLADCGY 608  
547 LQGLPREYVNAHCLPCHPECPQNGSVTCGPEADQCVACAHYKDPFVARGP----- 601

Db	609	ISNAIK--FDNRITCKIKCHECR-----TCNGAGADHCOECVHVDGGCHVSECKRKNY	660
Qy	602	-----SGVK-----	PDLSYMEIW 614
Db	661	DGVCRECHATGDCGTGPRDITIGIGACTTCNLAIINNDATVYKRLKXDKDCPD-GY--FW	717
Qy	615	KF--DDEGACQP-----CPI-----NCTH-----	632
Db	718	EYVHVEQGSLEKPLAGRAVCRKCHPLCELTINYGYHEOVCSKTHYKREOCETECPADH	777
Qy	633	-----SC-----VDLDDKG-----CPAQE 646	
Db	778	YTDGEQRGCFQHNREPCNGCTGPGADDKSCSRNFKFLPDANETGPRYVNSMFMNCTSKCPLEH	837
Qy	647	R-----ASPLTS-----IVSAVGLLVVLGVFGIILKRRQ 679	
Db	838	RHVNYQYTAIGPYCAASPPRSKITYANLDVNNIFITGAVALVPTICILCV--TYICRQK 895	
Qy	680	OKIRKYT--MRRLQETELVEPLTPSGAMPNQAORRIKETELRVKVLVIGSGAFETVYK 737	
Db	896	QKAKETVYKMTALSGCEDSEPLRNSNGANLCKIRIKYDAELRGVGLMGAFERYVK 955	
Qy	738	IWIPDGENVKIPVAIKVLRENTSPKANKEIDEAYVMAVGSPPVSRLLGICLTSTVOLV 797	
Db	956	VWVPEGENVKIPVAIKELIKSTGASSEPLREAYIVMSEEHVNLTKLLAVCMSSQWMLI 1015	
Qy	798	TOLMRYGCLLDHVRENRGRLSQDILLNMCQAKMSYLEDVRLVHRLAARLVLYKSPN 857	
Db	1016	TOLMPLGCLLDVYNNRRKIGSKALLNSTQIAKMSYLEERLTVHRLAARLVLYQTPS 1075	
Qy	858	HVKITDFGLARLLDIDETETVHADGSKVPIKMMALFSLRRPPTHOSDVMSGYTVWELMT 917	
Db	1076	LVKITDFGLAKLLSDSNERYKAAGKMEIKMLALECTINRRPFTSKSDVMAAGVHIWELT 1135	
Qy	918	FGAKPYDGIPAEPIPDLLIEKGERLPPICTIDVTVMVVKCMIDSECRPPREFIVSEFS 977	
Db	1136	FGOREHENIPAKDIPDLIEVGKLEQPEICSDIDCYTLSCWHLDAARPTFKQLTTFVA 1195	
Qy	978	RMARPORFVYIIONEDLG--PASPLDSFFVYSLLED--DMGDLVDAEYVLVYQQGFCF 1032	
Db	1196	EPARDPGYLLIPGDKFTRLRP-----YISQDKDLIRKLAPPTDSEAIARPDVYLQ 1248	
Qy	1033	PDPAAGAGMVNHHRRSSSTRSGGDLTLGLPESEEAR-----RSPLASSEGAGSPVF 1086	
Db	1249	PKAAPGFS-----HRTDCT-----DENPKLNRKYCKDPSNKNSSGTGDER 1287	
Qy	1087	DG---DLGMAKAKGLQSLPTHDPBQLQYSEDPYVLPSEFSDGYAPLITGSPPOREYVNP 1143	
Db	1288	DSSAREVGVGNLR-----LDLPVDDDDVLLMP--TCQPGNNNNNM 1325	
Qy	1144	DVRAQPPRPGRCPLPAARPAGATLERAKTLSPGKNGVGVKDVAFAGAVENEYV-----TP 1199	
Db	1326	N-----NPNQNMMAAVGVAAGYM-----DLIGPVASVDNPEYLLNQTL 1364	
Qy	1200	QCGAAPPOH-----PPPAFSP-AFDNLYYWD 1224	
Db	1365	GVGSESPITQTIGIPVMGGPGTMEVYKVPMPSEPTSPDHEYND 1408	
RESULT 12			
ERBB_ALV	1	STANDARD;	PRT; 634 AA.
AC	P00534;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).		
GN	V-ERBB		
OS	Avian leukosis virus.		
OC	Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.		
OX	NCBI_TaxID=11864;		
NN	[1]		

```

RP SEQUENCE FROM N.A.
DR MEDLINE=85228222; PubMed=298784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rotman F.M.,
RA Critenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroleukosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AYIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M1006; AAA48763.1; ALT_INTT.
DR PIR: A00643; TYCHLV.
DR PIR: B00643; TYFVLV.
DR HSSP: P1362; IFGK.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase; 1
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KM Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E3JA0B0D1FCC CRC64;

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 3,4e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

OY 587 CAHYKDPFCVCARCPGVKPDLSYMIWKPFDEEGACQPCINCTHSQVLDLDCGCPAQ 646
DB 3 CAHFIDSPHCVCAPVAGLNDTL-VMKYADANAVALCOLCHPNCRTCKGPGLEGCP--- 58
OY 647 RASPLTSIVAVV-GILVVVLGVVFGILIKRROKTRKTRMRLDLDETVLVEPLTSGA 705
DB 59 NGSTPTSIAGVVGGLLCLVVGIGLVLRR-HIVRKTLRRLDERELVEPLTSPGE 117
OY 706 MPNOQMRILKETELRYKVLGSGAFGVYKGIWIPGENVKIPVALKVLRENTSPANK 765
DB 118 APNQHILRIKETEFKVKVGLSGAFGVYKGLMIPGEKVKIPVALKEIREATSPANK 177
OY 766 EILDEAVYMGVSGPYVSRLLGICLTSTVQLVTQMLPFYGLLDVHRENRGRSGODLLN 825
DB 178 EILDEAVYMASVDPNPHVCRLLGICLTSTVQLTQMLPFYGLLDVIREHKONIGSYLLN 237
OY 826 CMQIAKMSVLEDEVRLVHRDLAARNVLYKSPNHYKITDFGLARLLIDETFYHADGGRVP 885
DB 238 CVQIAKMSVLEDEVRLVHRDLAARNVLYKTPHVKITDFGLARLLIDGDEKVEYHAEGRVP 297
OY 886 IKMMALLESILRRRFTHSDVMSYGVYTWELMTGCAKRYDGIIPARETIDILEKGRPLPP 945

```



```

Db      298 IKMMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy      946 ICTIDVYMWKCMIMIDSECRPRELVSFSSRAARPORFVYQ-NEDGAPAPLDSTF 1004
Db      358 ICTIDVYMWKCMIMIDADSRPKRELIAFSKKAARPPRLVYIQGBRNHLPSTDSKF 417
Qy      1005 YRSLLDDMDGLVDAAEYLVPQGFPCPPAPGAGVNHHRHSSSTRSGGDLTLGLE 1064
Db      418 YRTLMEEDMEDIVDADEYLVPHGFF-----NSPST----- 449
Qy      1065 PSEEBARSP-----APSEGASDVDFDGLGMAAGLQSLPTHDPSPIQRYSEDPTVP 1119
Db      450 -----SRTPLSLSLSATSNNSATNCID-----RNGQGHPRVEDSFQRYSSDPTGN 495
Qy      1120 LPSET--DGVVAPLTCSPQPEYVNPQVRPQPSRPGRLPAAAPAGATLERATLSPGK 1177
Db      496 FLEESIDDFL-----PABEYVNO--LMPKKSTAN-----TMVQ 526
Qy      1178 NGVVKDF-----AFGAVENPEYLTPOGGAAPQHPAPAPAFDNLVY 1222
Db      527 NQIYNISLTAISKLPMDSRQNSHTADVNDPEYL-----NTNQSPLAKTVFESSPY 578
Qy      1223 WDO-----DPE-----RGAPSTFKGTPTAENPEYLGIDVP 1254
Db      579 WISQGNHQLNDPNDYQODFLPNETKPNGLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 01, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain Es4).
OX Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
OX RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuit B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "sequencing the erbB gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC - MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: K02006; AAA42394.1; ALT_INIT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A06544; TVYU.
DR HSSP: P11362; IEGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE: Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146
FT BINDING 165 165
FT ACT_SITE 257 257
FT CONFLICT 140 140
FT CONFLICT 146 146
FT SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

```

Query Match 25.0%; Score 1703; DB 1; Length 604;  
 Best Local Similarity 52.2%; Pred. No. 9.6e-84;  
 Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```

Qy 587 CAHYKDPFCVAPRPGSVKRDLSYPIWKRPDEGACOPCEINCTHSCVDLDDGCPAQ 646
Db 3 CAHFTDGPCHVCVACPAGVLGENDTL-VKRYADANAVQULCHPNTCTCKCGGLGCP-- 58
Qy 647 RASPLTSTIVSAV-GILLVVVLGVFGILIKRQOKIRKYMRLLOETELVEPLTPSGA 705
Db 59 NGSTPTSIAGVGVGGLCLVVGAGIGLYLRR-HYRKRTLRLLDBRELVEPLTPSGE 117
Qy 706 MPNQAQRILKTELRLKRVKVLGSGAGFTVYKGIWIPGSENVKIPVAIKVIRENTSPANK 765
Db 118 APNQAHLRIKETEFKKVYKVLGSGAFGTIYKGLIPGEKVKIPVAIKELREATSPANK 177
Qy 766 EILDEAVVMAGVSPVYSLIGTCTSTQVLVQMLMPEGCLDVIREHKDNGSYLLNW 825
Db 178 EILDEAVVMASVNDPVCRLIGICTSTQVLVQMLMPEGCLDVIREHKDNGSYLLNW 237
Qy 826 CMOIAKMSYLEDRLVHRLAARNVLYKSPNHKRTDPSGLARLLDDETYHADGGVP 885
Db 238 CVOIAKAMNTLEERLVRDLAARNVLYKTPQHKITDFGLAKLGADKEYHAEGKVP 297
Qy 886 IKMMALESILRRFTTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQP 945
Db 298 IKMMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYMWKCMIMIDSECRPRELVSFSSRAARPORFVYQ-NEDGAPAPLDSTF 1004
Db 358 ICTIDVYMWKCMIMIDADSRPKRELIAFSKKAARPPRLVYIQGBRNHLPSTDSKF 417
Qy 1005 YRSLLDDMDGLVDAAEYLVPQGFPCPPAPGAGVNHHRHSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDADEYLVPHGFF-----NSPST----- 449
Qy 1065 PSEEBARSP-----APSEGASDVDFDGLGMAAGLQSLPTHDPSPIQRYSEDPTVP 1119
Db 450 -----SRTPLSLSLSATSNNSATNCID-----RNGQGHPRVEDSFQRYSSDPTGN 495
Qy 1120 LPSET--DGVVAPLTCSPQPEYVNPQVRPQPSRPGRLPAAAPAGATLERATLSPGK 1177
Db 496 FLEESIDDFL-----PABEYVNO--LMPKKSTAN-----TMVQ 524
Qy 1178 NGVVKDF-----AFGAVENPEYLTPOGGAAPQHPAPAPAFDNLVY 1222
Db 525 --VQNGIYNISLTAISKLPMDSRQNSHTADVNDPEYL-----NTNQSPLAKTVFESSPY 574
Qy 1219 NLYWDDPPERKAPSTFKGTPTAENPEY 1248

```



Dd		178	EILDEAYVMAVDNPHVRCLLGICLTSTVOLITQLMPRGCLLDITYREKHNDIGSQYLLNM	237
Oy		826	CMOJAKGMSYLEDEVRLVARDLAARNVLKVS PNHVKYITDFGLARLLDIDETEHADGKKVP	885
Dd		238	CVQJAKGNMYLLEERHMHVARDLAARNVLVKTPOHVKITDFGLAKQGADEKEYHAEGKKVP	297
Oy		886	IKMMALESILARRRFLHQSDWVS YGYTWTMLTFGAKPRPDGIPAREIPLDLEKGERLPPOP	945
Dd		298	IKMMALESILHRRIYTHQSDWVS YGYTWTMLTFGSKPYDGIPASISSVLEKGERLPPOP	357
Oy		946	ICTIDVVMIMKCMWNIDSECPREFELVSEFSRMARDQRFRVIO-QNEDLGAPSLDSRF	100-
Dd		358	ICTIDVVMIMKCMWSDDSRPKPELFELAEFSSKQARDPRRYLVLIQGDSRMHLPSPTDSKF	417
Oy		1005	YRSILEDMDMKDLVDAAEEYLVPQOGFFCPPDPAPGAGVHHRRHSSTRSGGDILTGLE	106
Dd		418	YRTLMEEEDMEDDIYADDEXLYVPHGCF-----NSPST-----	449
Oy		1065	PSEEBAPSPV-----APSEGAGSDVFDDGLGMGAUKGLQSLPTHDPSPLORYSEDPTVP	111
Dd		450	-----SRTPLLSSLATSATSNNSATNCIDRNGC-----H-----	476
Oy		1120	LPSETDGVVALTCSPOPEYNQODVVRQPQSPREBGP.LPAARPACAT-LERAKTLSPCKN	117
Dd		477	-PVREDGGL-----PAEYVNO--LMPKKKPSTAMVNQNIYYSILTAISKLPIDSRYON	527
Oy		1179	GVMKVDFAFPGAVENPEYL 1197	
Dd		528	-----SHSTAVDNPETYL 539	
 RESULT 15 EGFR_CHICK				
ID	_EGFR_CHICK	STANDARD:	PRT:	703 AA.
Ac	P13387;			
Dt	01-JAN-1990 (Rel. 13, Created)			
Dt	01-JAN-1990 (Rel. 13, Last sequence update)			
Dt	15-JUN-2002 (Rel. 41, Last annotation update)			
De	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER			
De	(Fragment).			
Gn	EGFR.			
Os	Gallus gallus (Chicken).			
Oc	Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;			
Oc	Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianine;			
Oc	Gallus.			
Ox	NCBI_TaxID=9031;			
Rn	[1]			
Rp	SEQUENCE FROM N.A.			
Rx	MEDLINE=88261272; PubMed=3260329;			
Ra	Lax I., Johnson A., Howk R., Sap J., Bellac F., Winkler M.,			
Ra	Ulrich A., Vennstrom B., Schlessinger J., Givol D.,			
Rt	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,			
Rt	expression in mouse cells, and differential binding of EGF and			
Rt	transforming growth factor alpha."			
Rl	Mol. Cell. Biol. 8:1970-1978(1988).			
Cc	-I- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,			
Cc	AND ALSO OF TGF-ALPHA, AMPHIPEGULIN, HEPARIN-BINDING EGF, GP30 AND			
Cc	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).			
Cc	-I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
Cc	tyrosine phosphate.			
Cc	-I- SUBCELLULAR LOCATION: Type I membrane protein.			
Cc	-I- MISCELLANEOUS: Binding of EGF to the receptor leads to			
Cc	dimerization, internalization of the EGF-receptor complex,			
Cc	induction of the tyrosine kinase activity, stimulation of cell DNA			
Cc	synthesis, and cell proliferation.			
Cc	-I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.			
Cc	-----			
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration			
Cc	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
Cc	the European Bioinformatics Institute. There are no restrictions on its			
Cc	use by non-profit institutions as long as its content is in no way			
Cc	modified and this statement is not removed. Usage by and for commercial			
Cc	entities requires a license agreement (See <a href="http://www.ebi.ac.uk/ncbi/pubmed/misc/copyright.html">http://www.ebi.ac.uk/ncbi/pubmed/misc/copyright.html</a> )			

or send an email to [license@sib-ch](mailto:license@sib-ch).

```
CC -----
DR EMBL; M20386; AAA48760.1; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; Fu; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 1 30
FT DOMAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT TRANSMEM 31 654 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 655 667 POTENTIAL.
FT DISULFID 668 667 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 703 703
SQ SEQUENCE 703 AA; 77427 MM; AFE2DE11B735A690 CRC64;
```

Query Match 23.6%; Score 1610; DB 1; Length 703;  
Best Local Similarity 44.4%; Pred. No. 1e-78;  
Matches 314; Conservative 113; Mismatches 252; Indels 28; Gaps 12;

```
QY 8 RMGLLALALPPGAA-----STQVCTGTDKMLRLPASPETHLDMRLHLYOGCCOVQGNLE 61
DB 13 RGAALVLLGLGVALGSAVEEKVCGGTNNKLTQLGHVEDEFTSLQRMYNNECVLNLNLE 72
QY 62 LTYLPTNASLFLQDIQEVGVYLIANOVROVPLQRLIVRGTLQFEDNYALAVLDNGD 121
DB 73 ILYVENHRDLTLFKTIQEVAGVYLIALNWDVIPLENLQILIRGNVLYDNSFALAVALSNYH 132
QY 122 PLANTTPVTGASGGLRELQRLSTELIKGCVLIQRNPOLCYODTIIMKQVIKANSKFI 181
DB 133 -MNAKTO-----GARELPMKRLSELINGVXISNNPKLCNMDTVLAMD-IIDTSRKL 182
QY 182 GITEI-LNRSBACHPCSPMCKSGSRGWESSEDCOSLTRVCAGCA-RCKGRLPTDCCHQ 239
```

```
DB 183 TVLDPASNLSSCPKCHPNCTEDHCGAGQONCQTLTRVITCAQDCSGRCRGVPSDCCHQ 242
QY 240 CAAGCTGPKHSDCLACGHPHNSGI CELHCPALVYNTDTRESMNPGRYTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACKKPRDDATCTKTPPLVLYNPOTVDVNPBGKTSFGATCVREC 302
QY 300 PNYVLSTDVGSCTLVCPILHNOEVTAEADGTORCEKSPCARVCYGLGMEHLREVAVTSA 359
DB 303 PHNYVTVDHGSCVRSCTVDYEV-EENGVAKCKKCDGLCSRYNCIGIGELKGLISINAT 361
QY 360 NIOEPACCKKI FGSIAFLPESPDGPASNTAPLOPEOLQVFETLEEITGVLYISAMPDSL 419
DB 362 NISFKXCTKINGVSIPLPAFLGDAFTKTLPLDPKKLDVFRVKEISGFLLIQAMPDNA 421
QY 420 PDLVSFONLQVIRGRIILHNGAVSLTLOGIGISWIGRLSRLSGLALIHINTHLCVHT 479
DB 422 TDLYAFENLEIRKRYKHQVSLAVNLKIQSLGLSLKEISDGLAIMKNKLCAVDT 481
QY 480 VPMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPPTQVCVNSQFLRGQE 539
DB 482 MNMRSLFATQSOKTKIIONNNKNDCTADRVVCDPLCSDVCGMGPGPHCFSCRFSSQKE 541
QY 540 CVEECRVLOGLPREYVNAHCLPCHPECPONG---SVTCFGEPADOCVACAHYKDPPEC 596
DB 542 CVKOCNIIQGEPEFERDSKCLPCHSECLVQNSTAVNTCSGPGPDHCKMCAHFIDGPHC 601
QY 597 VAPCGSVKPDLSMPYIKFPEDEGACOPCPINTHSCVDLDDKCPAEQASPLTSTVS 656
DB 602 VKACPGAVLGEENDTL-VMKYADANAVCOLCHPNCTRGCKPGLEGCP---NGSKTFSIAA 657
QY 657 AVV-GILLVVVLGVFGILIKRROQKIRKXTMRLLQETELVEPLTP 702
DB 658 GVGGLGLLVVGLIGILYLRH-HYRKKTIRLLQERELVEPLTP 703
```

Search completed: July 22, 2003, 08:44:44  
Job time : 20.2304 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-149-163-12

Sequence: 1 METALCRMGLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table:

BLOSUM62

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_TREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phase:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_rvivirus:\*  
17: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6225	91.4	1259	6	018735
2	3165	46.5	1209	11	090X70
3	3133	46.0	1210	11	09EP98
4	2734	40.1	1165	13	09YH40
5	2705.5	39.7	1137	13	09W6F6
6	2317	34.0	1328	4	P79754
7	2032.5	29.8	1433	5	09BIH9
8	1802.5	26.5	419	4	09UK79
9	1739	25.5	367	11	08RX1
10	1720	25.3	729	15	086712
11	1718	25.2	567	15	086714
12	1697.5	24.9	412	4	08WYV0
13	1653.5	24.3	962	15	064895
14	1645	24.2	545	15	085468
15	1524.5	22.4	655	11	09WYF5
16	1508.5	22.2	643	11	09ERV6

17	1276	18.7	1193	5	09Y1X8	09Y1X8 ephradya f
18	1194.5	17.5	1368	5	023821	023821 caenorhabdi
19	1162	17.1	1717	5	026566	026566 schistosoma
20	1141	16.8	527	13	090836	090836 gallus gall
21	1019.5	15.0	478	11	09ESE0	09ESE0 rattus norv
22	956.5	14.0	599	13	09ESH2	09ESH2 rattus norv
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.0	176	11	0923V5	0923V5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	08SZW1	08SZW1 dirosophila
27	762.5	11.2	1362	13	09PYZ4	09PYZ4 xenopus lae
28	754.5	11.1	311	13	09Y162	09Y162 xiphophorus
29	737	10.8	1671	5	09NUV5	09NUV5 biophthalari
30	726	10.7	331	4	09BUD7	09BUD7 homo sapien
31	712.5	10.5	1368	13	08UM85	08UM85 paracichthy
32	706	10.4	149	6	09BG66	09BG66 oryctolagus
33	697	10.2	1418	13	093457	093457 scophthalmu
34	689.5	10.1	1369	13	08UM86	08UM86 paracichthy
35	679	10.0	1358	13	073798	073798 bombyx mori
36	663.5	9.7	1472	5	09U5A8	09U5A8 paracichthy
37	659	9.7	1412	13	08UM84	08UM84 paracichthy
38	648	9.5	1245	13	09YGH8	09YGH8 scophthalmu
39	643.5	9.4	1418	13	08UM83	08UM83 paracichthy
40	639	9.4	1371	11	09QVW4	09QVW4 rattus sp.
41	617.5	9.1	2144	5	09VD94	09VD94 dirosophila
42	588.5	8.6	987	11	091YMO	091YMO mus musculu
43	587.5	8.6	1036	4	007912	007912 homo sapien
44	587.5	8.6	1055	11	054967	054967 mus musculu
45	586.5	8.6	1091	4	09UMQ4	09UMQ4 homo sapien

## ALIGNMENTS

RESULT 1  
ID 018735 PRELIMINARY; PRT; 1259 AA.  
AC 018735;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE EDB-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "cDNA cloning of erdb-2 from canine mammary gland."  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1; -.  
DR HSP; P11362; IFC.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_Pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP\_2.  
DR Pfam; PF000001; Euk\_Pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PSS0011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PSS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;



Db	15	LALALCAG-----GALBKKVVOCGSINRLTOLGTFEDHFLSLQRPFNCEVVLGNLE	66
OY	62	LTYETNASLSFLODIOEVGVYLI,IAHNOVRQVLPQRLRIYRGTOJ,FEDNYVALVLDNG	121
Db	67	ITYVQRNVDLSPLKLTIOEVAGVYLI,ALMTVERI+PLENIQIIRGNALYENTYALAVLSN--	124
OY	122	PLANTTPYTGASPGGLRELOJLSLTLEIKGCVL,IOIRNPOLCYQDITPMKDOYIKANSKFI	181
Db	125	-----YGNKTKOLRELPMNLOEIIIGAVRFSNNPILCMETIORMD---IVQDFVL	173
OY	182	GITELNRSR---ACHRPSMCKSGN,CWESSSDQSLTRTVAGCA--RCKPRLPTDCH	237
Db	174	SMSMDVORHLTGCKCDSPSCNGSCWGBEENCKLTKITICAQCCSRCKRSPSDCH	233
OY	238	EQACAGCTGPKASDCLACLFHNSGISCELHCALVTYNTDTEESMPNEGRTYFASCVT	297
Db	234	NQACAGCTGPRSDCLVCHRFDEATCKDTCGERPLMYNPTYQMDVNBGAKTSFATCYK	293
OY	298	ACPNYLSTDVGSCTLVCPRLHNOEYTABDGTORCEKSKPCARVCYGLAMEHREAVT	357
Db	294	KCPRYVYVTBHGSCYRACGPDYEEV-EEBGVSKCKCKCGPCRCVKNCGIGIGEFKOTLSIN	352
OY	358	SANIOEFAGCKKIFESLAF,FPESPQGDPAASNAP,OPQOLOVFETLEETIGLYISAMPD	417
Db	353	ATNIGHFXYCTAISDLDHLI,LPVAFGDSFTRRP,LPDRELEETLTKVKEITGGLLQAMPE	412
OY	418	SLPDLISVQNLQVINGRILHNGVSLTLOGLISWIGLSRLSELSSGLAI,HHNTHLCFV	477
Db	413	NMTDLHAENLEIIRGRTRKQHQFSLAVVGLNITSGLRSLEIDGVIYISGNNNLCYA	472
OY	478	HTVPMDOLFRNDHQALHTANR,PEDECVGEGGLACHQOLCARGH,CWGPBPPTOCVNSQOPLRG	537
Db	473	NTIIMWKLFEPNQTKT,IMNNRAEKDCKATNHNCP,LCSSBEGCWGPBPPTDVCSCQNSRG	532
OY	538	QECVEECGVLOGLPREYVNAHCLFCHHECQONQNSVTCFGEBAQCVACAHYKRRPCV	597
Db	533	RECVDCNKILGEPEFEFENSECICQCHBCLPQTNWITCTGGRPNCKICAHYVYDGPCHV	592
OY	598	ARCPSGVPRDLSYMPIMKFPDEEGACQPCPICTHSCVCDLDDKCGPAEORASP-LTISVS	656
Db	593	KTCBSGIMGENNTL-VMKFADANNVCHLCHANCTYGCAGPGLKGC--QOPBSPKIPSTAT	649
OY	657	AVVGLILVVVLGVVFGI-LIKRQOKIRKYTRMLLOETELVEPLTPSGAMENOAKMIL	715
Db	650	GIVGGLLFIIV-VALGIGLFMRRRQOLVKKRTRLRLQRELYEPLTPSGEABNQHLIL	708
OY	716	KETELRKRV,GSAGFTVYKGIW,PDGENVKI,PAIATVLRNTSPKANKELIDBAYYMA	775
Db	709	KETERKKIKV,GSAGFTVYKGLW,PEBEKVKI,PAIKELREATSPKANKELIDBAYYMA	768
OY	776	GVGSPYVYRLGICIT,STVYOLVTO,MPGCLLDHYENRGRLGSGDILLNMCQIAKGSY	835
Db	769	SVDPNPHVRLGICIT,STVYOLITQ,MPGCLLDYREHNDNGSYLLNMCQIAKGMNY	828
OY	836	LEDVRLVHRD,LAARNVLKVS,PNHVKITD,FGIARL,LDIDETEHADGKVP,IKMMALESIL	895
Db	829	LEDRLRLVHRD,LAARNVLKTPQHVKITD,FGIARL,KGAEKEYNHAGKVP,IKMMALESIL	888
OY	896	RRRFTHGSDDWSYGTTWELMT,FGAKPPDGI,PARBI,PDLEKEGEFL,POPPICTIDVYIM	955
Db	889	HRITTHGSDDWSYGTTWELMT,FGSKPDDGI,PAISEI,SIILEGEBL,POPPICTIDVYIM	948
OY	956	VKCMNIDBECR,PRFELVSEFSRMA,RPORFV,IO-NDLDGPASPLDSTF,RSILEDDM	1014
Db	949	VKCMNIDBDS,PKRELEILFESK,AKARDQ,RYLIVIGDDBMIL,PSPTSDNPF,RYALMEEDDM	1008
OY	1015	GDVLDAEBEYLV,POQGF,FCPDBA,PGAGM,VNHRHRS,STRSGGCDLTGL,LESSEBEPSP	1074
Db	1009	EDVVDADAEYLV,POQGF-----NSPST-----SRTF	1034
OY	1075	LAPBEGAGSV,FDGDLGMAAKGLOSLF,PHRPSLP,OKRIS,EDPVPYLP,SBST--DGVAVLT	1132
Db	1035	LSSLSIASNSN-----SSTVACINRNGSCRVK,DKDAFLQ,RRSPTSSTVLT,TDNIDDTFL---	1086

```

Qy      1133 CSPOEYVNOGDVARPQPSPREBGLPARAPAGATLIERAKTLSSPCKNGCVKDVFACGAVE 1192
         |||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       1087 --PWPEYLINQ--SVPRRPAQGVONVEYNHNPILHP-----AFGRDLHYON--DHSNAVS 1133
         |||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy      1193 NPEYL-TEQGGAARQPPHPPAFSPAFDLLYYWDO-----DP-----PERGAPS 1235
         |||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       1134 NPETLINHQ-----PTLSGGFSDSSALMIKGSHMSLDNPDVQQDFPFKAEPNG 1184
         |||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy      1236 TFKGTPTANPEYLGLDVP 1254
         |||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       1185 IFKG-PFAENAEYLRVAVP 1202
         |||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

RESULT 3
QE9EP8          PRELIMINARY;        PRT;    1210 AA.
ID              QE9EP8
AC              QE9EP8;
DT              01-MAR-2001 (TREMBLrel. 16, Created)
DT              01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT              01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE              Epidermal growth factor receptor isoform 1.
GN              EGFR.
OS              Mus musculus (Mouse).
OC              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
RN              NCBI_TaxId=10090;
RP              SEQUENCE FROM N.A.
RC              STRAIN=C3H/10J, 129/SvJ, AND 129/SVEYTAC;
RA              Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA              Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
RA              Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA              Mahle N.J.;
RT              "Comparative genomic sequence analysis and isolation of human and
RT              mouse alternative Egfr transcripts encoding truncated receptor
RT              isoforms.";
RL              Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RM              [2]
RN              RN
RP              SEQUENCE FROM N.A.
RC              STRAIN=C57BL/6J;
RA              Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA              Schehl C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
RA              Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA              Mahle N.J.;
RT              "Comparative genomic sequence analysis and isolation of human and
RT              mouse alternative Egfr transcripts encoding truncated receptor
RT              isoforms.";
RL              Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RM              [2]
RN              RN
RP              SEQUENCE FROM N.A.
RC              STRAIN=C57BL/6J;
RA              Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA              Schehl C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
RA              Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA              Mahle N.J.;
RT              "Comparative genomic sequence analysis and isolation of human and
RT              mouse alternative Egfr transcripts encoding truncated receptor
RT              isoforms.";
RL              Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RM              [2]
RN              RL
RP              Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR              EMBL; AF275366; AACG28045.1; -.
DR              EMBL; AF275364; AACG28045.1; JOINED.
DR              EMBL; AF275365; AACG28045.1; JOINED.
DR              EMBL; AF275367; AACG24386.1; -.
DR              HSSP; P11362; IFGK.
DR              MGP; MG1.95294; Egfr.
DR              InterPro; IPRO00345; Cytc_heme_bind.
DR              InterPro; IPRO00494; EGFR_L_domain.
DR              InterPro; IPRO00719; Euk_pkinase.
DR              InterPro; IPRO02174; Furin-like.
DR              InterPro; IPRO02290; Ser_thr_kinase.
DR              InterPro; IPRO01245; Tyr_pkinase.
DR              Pfam; PF00757; Furin-like; 1.
DR              Pfam; PF00069; pkinase; 1.
DR              Pfam; PF01030; Recep_L_domain; 2.
DR              PRINTS; PR00109; TYRKINASE.
DR              ProDom; PD000001; Euk_pkinase; 1.
DR              SMART; SM00261; Fu_5.
DR              SMART; SM00220; STKC; 1.
DR              SMART; SM00219; TYTKC; 1.
DR              PROSITE; PS00190; CYTOCHROME_C_ UNKNOWN_1.
DR              PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
DR              PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR              PROSITE; PSS00109; PROTEIN_KINASE_TYR; 1.

```

KM ATP-binding; Receptor; Transferase.  
SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 46.0%; Score 3133; DB 11; Length 1210;  
Best Local Similarity 49.6%; Pred. No. 1.7e-227;  
Matches 629; Conservative 170; Mismatches 367; Indels 102; Gaps 21;

```

11 LLLALLPAGA--STGYCTGDMKRLRLPASPEHLLDMRLHLYGCGOVVQGNLELTLYPTN 68
14 LRLRLCAAGALEKKEKVCQGSNRLTQLGTFEDHFLSLQRMYNNGCEVVLGLLETTYORN 73
69 ASLSFLDDIOEVQGVLIHANQVROVPLQRLRIVRGTOLEFEDNALAVLNDGDLNMTTP 128
74 YDSFLKTIQEVAGVYLIALNTVERIPLENLQIIRGNALYENTYALALISN----- 124
129 VTGASPGGLRELRLSLTEILLKGVLIQORNPOLCYODITLLKDOYIRANSFGITELNR 188
125 -YGNRRLGLRELPRNRLQELIIGAVFSNNPILCNMDTIQWRDVIQVNFMSNMDLSH 183
189 SRACHPSPMCKSGRCMGESSEDCSLTRTVACAGCA-RCKGPIPTDCHEQCAAGCTGP 247
184 PSSCPKCDPSCPNSSCMGSGEENCKLTKIICAOQCSHRCGRSBSDCCHQCAAGCTGP 243
248 KHSDCLACLPHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCVTACPYNYLSTD 307
244 RESDCLVQCKQDEATCKDTCRPLMLYPTTYQMDVNEGKYSFGATCVKCKPRNYVYTD 303
308 VGSCTLVCPRLNQVTAEDGTQRCCKSPCARVCYGLGMHLEBRAVNTSANIQEFAGC 367
304 HGSQVRAQGPDYEV-EEDGIRKCKKCGPRKVCNGGIEFQDTLSINKTINKHFYKC 362
368 KKIIFSLAFLPESPDPASNTAPLOPELOVFEETLEETLYLYISAMPDSLPLSVYN 427
363 TAIQGDHLIPVAKKGDSFTTPRDLRELEILKTVKEITGFLLIQAMPDWTDLHAFEN 422
428 LQVIRGRILHNGAVSLTQGLGSLWGLRSLRELGSGLALIHNTHLCEFTVTPMDQLFR 487
423 LEIRGRTRKHQGSFLAVVGLNITSLGRSLKEISDGDVITSGNRNLCYANTIMWKILFG 482
488 NPHQALHTANRPDECEVGBGLACHQLCARHCHGPGTQCVNCSQFLRQGECEVECVL 547
483 TPNQKTKIMNRRARCKCAVNVHVCNPLCSSGCGWPEPRDVCSSVSRGSECEVKCNIL 542
548 QGLREYVYNAHCLPCHPECOFONGSVTCFPEADQCVACAHYKDPFCVARRCSGVNPD 607
543 EGEREFEYENSECICQHECELPQAMNITCTGRGPNICQCAHYIDGHCIVCTCAGINGE 602
608 LSYMPWKFPDEGACQPCPINCTHSCVDLDDKCGCPABQASPLTSIVSAVGLLVYL 667
603 NNTL-VMKYADANNVCHLCHANCTYGCAGPGLQGEVWPSGPKIPSIATGIVGGLFIV 661
668 GVVFGI-LIKRQOKIRKTYMRLLQETELVEPLTPSGAMNQOMRLKTELKRVYL 726
662 -VALGIGLPMRRRIIVRRRLRLQLQERLEVEPLPSEGAENQALRLILKTEFEKIKVL 720
727 GSGAFGVYKGIWIPDGENVKIPVALKYLRENTSPKAKEILDEAYYVAGSPVSRLL 786
721 GSGAFGVYKGLMIPBEGEKVPIPAIKELREATSPKAKEILDEAYYVWASVDNPHVCLL 780
787 GICLTSTVQLVTOIMPGCLLDHVENRNGRLGSDLLMWCQIAKMSYLEDVRLVHDL 846
781 GICLTSTVQLITLOLMPYCLLDVYREHKNIGSYQLLMWCQIAKMSYLEDRRLVHDL 840
847 AARVNLVKSPPVHVITDFGLARLLDIDETEHADGCKVPIKMMLESLIRRFTHQSVW 906
841 AARVNLVKTPOHVKITDFGLAKLLGAEEKETHABGKVPIMMALESLIYTHQSVW 900
907 SYGVTVMLMTFGAKPYDGIIPAREIPDLLEKGERLPPPCITIDVYIMVCMWIDSECR 966
901 SYGVTVMLMTFGSKPYDGIIPASDISISLEKGERLPPPCITIDVYIMVCMWIDNSR 960
967 PRFELVSEPSRMARDPQRFVVIQ-NEQLGASPLDSTFFYSLLDEDDMGLVDAEYLV 1025

```

```

Db 961 PKFRELLEFSKMARDPQRYLVLIQDERMLPSPPTDSNFYALNDEEDMEDVDAEYLT 1020
Qy 1026 PQGFFPCDPAPGAGVHHRRHSSSTRSGGDLTLGLEPSEEARPSPLASGAGSDV 1085
Db 1021 PQGFF-----NSPST-----SRTPLSLSATSN- 1045
Qy 1086 FDGDLGMAAKGLQSLPTHPSPLOQYSEDPVLPSET--DGYVAPLTCSPQPEYVNP 1143
Db 1046 ---NSTYACINRNSCKVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PVPEYVNO- 1095
Qy 1144 DVRPQPSPREGPIPAARPGATLERAKTUSPGKNGVVKQVFAFGAVENPEYL-TPQGG 1202
Db 1096 SVPRPAGSVQNPYPYHNQPLHP-----APGRDLHYQN--PHSNWGNPEYLYNTAQ-- 1143
Qy 1203 AARPHPRAFSPPAFDULYVWDQ-----DP-----PERGAPSTFKGTPTAENP 1246
Db 1144 -----PTCLSSGFNSPALWIKQSHOWSLNDPDYQDFFPKETKPGIKRG-PTAENA 1195
Qy 1247 EYLGADVP 1254
Db 1196 EYLRVAPR 1203

```

## RESULT 4

Q9YH40 PRELIMINARY; PRT; 1165 AA.

```

ID Q9YH40
AC Q9YH40,
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Winkler C., Welbrock C., Gomez A., Duschl J.,
RT Altschmidt J., Scharl M.;
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT overexpression and mutational alterations.";
RL Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Scharl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53471; AAD10500.2; -.
DR HSSP; P11362; IFCG.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.

```

SQ SEQUENCE 1165 AA; 129614 MW; 7F7EB3808771A74E CRC64;  
 Query Match 40.1%; Score 2734; DB 13; Length 1165;  
 Best Local Similarity 45.7%; Pred. No. 2,2e-197;  
 Matches 581; Conservative 165; Mismatches 386; Indels 140; Gaps 29;

```

QY 1 MELALCRWGLLALLPG-AAST---QVCTGDMKRLPASPETHLMDLRLHLYGCGCY 55
DB 4 LELLEL-----LLLLLISGRCCSTDPDRVCCTSTQMTM---LDNHLYLKKMKMGSGCNV 56
QY 56 VQGNLELYLTPINASLFLQDIQEVGYVLIHNOYROYPLQRLIRVGRQULFEDNYALA 115
DB 57 VLENELITTOENODLSFLOSIOEVGYVLIHNEVSTIPLVNLRLIRGNLVEGNFTLL 116
QY 116 VLDNGDPLNNTTPTVGASPGRLRELQRLSTELKSGVLIQRNPOLCYDITLMDKQYIK 175
DB 117 VMSNVQK-NPSSP--DVYQVGLKQQLQSLNLTLSGAVVSNHNPILCNVETIMMMDIVK 173
QY 176 ANSKFIGITELRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGC-ARCKGPLPTD 234
DB 174 TSNPTMNLIPHAFEROCOKCDPCVNGSCMAPGPGHOKFTKLLCAEQCNRCRCRGPID 233
QY 235 CCECCAGCTGPRKSDCLACTLHPNHSGLCELCPALVYNTDTFESMNPBERYTFGAS 294
DB 234 CNEHCAGGCTGPRATDCLACRDFNDGTCOTCPPKTYDIVSHQVVDNPNIKYTFGA 293
QY 295 CVTACPYNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARCYGLGMEHLREVR 354
DB 294 CKKECPNSNVYTE-GACVRSAGMLEV-D-ENGKRCRCPDGVCPVCCIGIGISLNTI 351
QY 355 AVTSANIOEFACCKIFGSLAFLPESFDGPPASNTAPLOPEOLQVETLEETGYLYISA 414
DB 352 AVNSTNIGSFNSCTKINGDIILNRNSFEEDPHYKIGPMPEHLMNLTVEITGYLYIWM 411
QY 415 WPDLSLPDSVFNLOVIRGRILHNGVYS-LTLQGLISWGLRSBELSGLALIHNTH 473
DB 412 WPEANTSSVFNQLEITIRKRTFSRGFSVVOVSHLQWGLSSLSLEVAGAVILNTPO 471
QY 474 LCFVATVPDQLFRNPQALHTANRPEDECVGEGLACHQLCARGHCMGPFPTQCVNCSQ 533
DB 472 LRYASTIMMRALFRSEDOGIEVDART-----ENQTCNNECSBDCMKPFGPTMCVSC 524
QY 534 FLRGECEVBECCVNLQGLPREYVNAHCLPCHPECCOPQNSVTCFGEADQVACAHYKDP 593
DB 525 VDRGRCAVASCNMLQGEPREAOVDGCVQCHOECIVQTSITLCYCGPANCSCAKAFQDG 584
QY 594 PFCVARCSPGVKPDLSYMPIMKFPDEBEGACOPCINCHTSQVLDLCKGPAQRASPLTS 653
DB 585 POCIRCPHGMIGDGDTL-IMKYADMGCGQCPCHQCTQCGSGPGLSGRGD-IVSHSSL 642
QY 654 IVSAVVGILLVVLGVFGLIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNQOAKR 713
DB 643 AVGLVSGLLITVALLIVLIRRRIRK-RKRTIRLLQEKLVLELTSGCAPNOAFIR 701
QY 714 ILKELTELKRVKVLGSGAFTYKGIWIDGENVKIPVAIKVIRENTSPRANKELIDEAVY 773
DB 702 ILKETEFKDRYLGSGAFGTIVKGLMNPGENIRIPVAIKVIRENTSPRANKELIDEAVY 761
QY 774 MAGVCSPVYSRLIGLCTSTVOLVQMLPBYGCLLDHVENRRLSODLLNMCQIAKAM 833
DB 762 MASVDHPVCRLLGLCLTSANVLQVLMFYGCLLDYVRQHORICQOMLLNMCVQIAKAM 821
QY 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDPGLABLLDDEFEYHADGKVPKIMMALES 893
DB 822 NYLEERHLVHRDLAARNVLKSPNHVKITDPGLSKLLTADEXGYAHGKVPKIMMALES 881
QY 894 ILRRFTQSDVWSGVTVMELMTFGAKPYDGIPIARETIDLEKEKRLPQPICTIDVYM 953
DB 882 ILQMTYTHQSDVWSGVTVMELMTFGSKPYDGIPIARETIDLEKEKRLPQPICTIEVYM 941
QY 954 IMVCKMIDSECRPFRELVSFSEMRARDPQRFVIVIONEDLGPASPLSTFRSLLEDD 1013
DB 942 IILCKMIDPSSRPFRELVEFSQMAADPSKIVIQG---NLSPSDRLRFLSSDD 998
  
```

QY 1014 MGDVLDAEELVYPQGFPCPDPAAGAWVHRRSRSTRSGGDLTLGLEPSEEAPRS 1073  
 DB 999 --DVVDADETL-----RYKXIN-RQGS-----E 1019

```

QY 1074 PLAPSEGAGSDVFDGLGMAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDYAPLT 1132
DB 1020 PCIPRPNGH-----PVRENSIALRYISDPTQNALFEDLDCH----- 1054
QY 1133 CSPQPEYVNOPDVRPP-----PSPRE-----GLLP-AARPAGATLEPAKTLSPCKNG 1179
DB 1055 -----BYVNOPGETSSRSLSDIYNVYEDLTDGCMGPVLSLSQAEITNFSREYLNTQNS 1109
QY 1180 VWKDVAFGAAVENPEYLTPOGGAAPQHPHPAFSPAFDULYYWDDOPPERGAPSPSTFKG 1239
DB 1110 L---PLVSSGSMDDPPY---QAG-----YQAAF-----LPQTGALTGNGMF 1144
QY 1240 TPTAENPEYIGL 1251
DB 1145 LPAENLELYIGL 1156
  
```

RESULT 5  
 Q9W6F6 PRELIMINARY; PRT; 1137 AA.  
 AC Q9W6F6;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase (Fragment).  
 GN ERB4.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDRAIN;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.;  
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in  
 RT embryonic chick hindbrain".  
 RL Mol. Cell. Neurosci. 13:237-258(1999).  
 DR EMBL; AF121963; AAD31764.1; -.  
 DR HSSP; P13562; IFGK.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001368; TNFR\_C6.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 1.  
 DR Pfam; PF02757; YLP; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TYKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00652; TNFR\_NGFR\_1; UNKNOWNV\_1.  
 KW Kinase; Tyrosine-protein kinase.  
 FT NON\_TER  
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 39.7%; Score 2705.5; DB 13; Length 1137;  
 Best Local Similarity 47.0%; Pred. No. 3e-195;  
 Matches 535; Conservative 170; Mismatches 357; Indels 77; Gaps 24;

```

QY 161 LCYQPTILMKQYIKANSKFIGITELNRSRACHPCSPCKSGRCWGESSEDCQSLTRTV 220
  
```

```
DB      3 LCFADTIIHQDIVRNPMWASFTLVPNTGSSGCGRCHKSCTG--RCMGPTENHCQTLLTKTVC 61
QY      221 AGGC-ARCKGLPTDCCBOCAAGCTGPKKSDCLACAFHNSGICELHCPALVTYNTDTF 279
DB      62 AEQDGCCTGYPVSDCCCHREAGCGSPKOTDCCACNNFDSGACTGQCPTFEYVNPPTTF 121
QY      280 ESMNPBGRYTFGASCVTACPYNYLSTDVGSCTLVCPILHNOEYVAEDGTORCEKSKRCA 339
DB      122 QLEHNNAKTYGAFVCYKCKPHNFV--VDSSSVACACSSKMEV--EENIGMKCKCTDTCR 179
QY      340 RVCYGLAMEHLREYVAITSANIOEFACCKIFGSLAFIPESFDCCPASNTAPLOEOLQV 399
DB      180 KACDGICTGSLVSAQIVDSSNIDFICTKINGNLIFLVGIGHDPHTHTAIIPEKINI 239
QY      400 FETLEEITGYLYISAMPDSL.PDLSVFONLQVIRGRIIHNNAVSLTLOGGLISWGLSLR 459
DB      240 FQTVREITGYLNTQSWPENNTDFRVFSNLVTIGRALYSGISLLILKQGGITSQFOSLK 299
QY      460 ELGSGGLALIHNTLFCVHTVPMQDLFRNPHQALLHTANRPEDECVESGLACHOLCARGH 519
DB      300 QISAGNIYITDNSLGYHTYVNTSLSTFSQKTVIHRNKKAENCTADGVCNCLSSDG 359
QY      520 CWGPGPTQCVNCSQFLRGQCEVGRVLOGLPREYVNAHCLPCHPEQCP--QNGSVTCFG 578
DB      360 CWGPGPQCLSCRFIRGTCIESCNLYDFEFREFANGSVCMEDCPQCEKMEKMTICYG 419
QY      579 PEADQVACAHYDOPFCVACRSPGVK.PDLSYMPIMKFPDEGACQCPICINCHSCVDLD 638
DB      420 PGPDHCKCFHFGKGRPCVEKCPDGLQGANSF--IFKXADRECHPHPCITGCGCPA 477
QY      639 DKGC-----PAORASPLTISVAVV--GILLVVVLGVVFGILIKRQCKIRKYT 686
DB      478 SHDCIYVPMTRQSTLPQHAR--TFL--IAAGVIGLFIIVIMGLFAYVVRKRSIK--KKRA 533
QY      687 MRLLQSTELVEPLTBSGAMPNOQMILKETEIRKTKVUSGAFGYTYKICM.PDGENV 746
DB      534 LRRL-ETELVEPLTBSGAMPNOQLRIKETEIRKTKVUSGAFGYTYKICM.PDGETV 592
QY      747 KIPALIVLBENTSPKANKELIDEAVMAGVSPYVRILGICTSTVOLTOLMPGCL 806
DB      593 KIPALIKILNETTGPKANVEFMDALIMASHDHLVRLGVCSPITQLTQOLMPGCL 652
QY      807 LDHVRNCRGLSGODLLNMCQIAKGSYLEVRLVHRDLAARNVLKSPNHVKITPGL 866
DB      653 LDYVHEKDNIGSOLLNMCVQIAKGMVYLEERLVRDLAARNVLKSPNHVKITPGL 712
QY      867 ARLLDIDETEHADGCVPIKMMALLESILRRFTHSDVWSYGVTVWELMTFGAKPYDGI 926
DB      713 ARLEGEKEKYNMGKMPKMMAL.ECIIHYRKFTHSDVWSYGVTVIWMELMTFGCKPYDGI 772
QY      927 PAREIPDLLEKGERLPQPICTIDVYIMVCMIDSECRPRPRELVSSEFSRMARDQRF 986
DB      773 PTEIIPDLLEKGERLPQPICTIDVYIMVCMIDSECRPRPRELVSSEFSRMARDQRF 832
QY      987 VVIQNEB--LGPASPLDSTFYRSLLEDMDMDLVDAAEYLVPOQGFPCDPAPAGGVNH 1045
DB      833 LVIQGDDRMKLPSPNDSKFFONLLDEBDLEDMDMAEEYLV--QAFNIPPIYLSRTIDS 891
QY      1046 RHSSSTRSGGDLITLLESEEARPS--PLAF--SEGASGVVDGDLGMAAGLOSLP 1102
DB      892 NRNOFYVRDGYAAEGGV--PMYRPAFCIIIEAIVAGATAEIREDCCNGTJLRKQVATL 950
QY      1103 THDPSPLOKSEDPVPLPS-----ETDGYVALPLCSPOPEVYVNO.PDVAPQPPREG 1155
DB      951 AKEDSSQKRSADPTVFIPEKVRIGELDEDCYMTFMDKPKTIDLVNVEENPFVSRKNG 1010
QY      1156 PLPA--RPAQATLERAKTILSPKNGVVKVVF-----AFGAIVENPEYLTPOGAAPQ 1206
DB      1011 DLQAVDNPETHN-----APNGQPKADEYVNEPLVNTFANTLENAEYL-----K 1055
QY      1207 PHPPAPSPAFDNLVYWDQDQPERGA--PSTFKGTPT-----AENPEYL 1249
```

```
DB      1056 NNLEPKAKKAFDNPDYNNHSLPPRSTLQHPDYLOESTKYFYKONGRIPIVAENPEYL 1114
RESULT 6
P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
EdbB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes.";
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSBP; P11362; IFGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Dkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-1like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Prodom; PD000001; Euk_Dkinase; 1.
DR SMART; SM00261; Fu; 3.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KM ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match 34.0%; Score 2317; DB 13; Length 1328;
Best Local Similarity 40.9%; Pred. No. 8.2e-166;
Matches 526; Conservative 153; Mismatches 417; Indels 190; Gaps 32;

DB      9 WGLLALLPP--GAATQ-----VCTGDMKLRLPASBETHLMDLRHLVYOGGVQGNIEL 62
DB      4 WRLLKCVASRLRASSQTOEAVCPGTONGLSSTGSEQENYLNKDKRYKGEIIMGNLEI 63
QY      63 TYLPTNLSLFLDIOEVGVYLIANQVROVPLQRLIRIVRGTOLFEDNVALAVLDNDP 122
DB      64 TQIESMNDPFLKTRREVTVGLIANNHFOEIRPLGQRLVRIGNSLYERRPALSFLN--- 120
QY      123 LNNTPVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILMKDQYIKANSKFIG 182
DB      121 -----YPRDG--PSGLNGLGMINLLEIIDGGVQIINNRYLRGVPVWRD--IIRNDAPIE 173
QY      183 ITELNRRACHPSPMKSGRCRWGSESESDQSLTRTYCAGGC--ARCKGLPTDCCBOCA 241
DB      174 I--QNGRGRV--CHKSC--GVYCWGPGKDCQOILTKTYCAQCNDRCGTSPRDCHIECA 229
QY      242 AGCGPAGSDCLACIHFHNSGICELHCPALVTYNTDFFESMPNBEGRYTFGASCVTACPY 301
DB      230 AGCKGRLDTCQFCARLFPNDGACVCPQCPQTLINYKQTFQMETNPNATKYQGISVCQCP 289
QY      302 NYLSTDVGSCTLVCPILHNOEYVAEDGTOR--CEKSKPCARVCYGLGMEHLREYVAITSAN 360
DB      290 HFV--VDSSSVCSVCPDKMEV--ERGSQRCCLCSGLCPKVCBGTGAE---QRTVDSN 343
QY      361 IQEPAGCKKIFGSLAFIPESFDCCPASNTAPLOEOLQVETLEITGYLYISWPSLP 420
DB      344 IDSPINCTKIQGSILHFLVTGILDDDFKNVPLDAKKLEVFRTVAEITDILNIOSWPKELN 403
```



```

OY /421 DLSVFONLQVIRGRILHNGAVSLTLOGIGISWILGRSLREUGSLALIHNTHTLCEVHTV 480
DB DLSVFSSLTITIOGSLFRGRSLMWRIPTLTSLGRSLREISDSVYISQNAHLCYHNTV 463
OY 481 PMDLFRNPH-QALLHTANREDECVEGEGLAGHOLCARGHMGCPPTQCVACOSOFIRQOE 539
DB 464 NMTOLEFRGRVRANSLNSNRPMACVADGRVCDPLCSGSGMGCPDGLCSCRYNRSGT 523
OY 540 CVEECRLVQLPREYVNAH-CLDPHPCOPONGSVTCFGEADQVCAHAYKDPFECVA 598
DB 524 CVAGCHNSGIPREFAGLNGVCAVCHPECKPQTGKASCTGGADECMACCTKFRDGPYMS 583
OY 599 RCPSGVPRDLSYMIMKPRDEEGACQCPRICTHSQVLDLDDKCGPABRASPILSYSAV 658
DB 584 SCFAGVA-DEBKGLIFKFPNREGHCEPHONCTGCGSGPGLNDC---LEARLTISGQI 639
OY 659 VGLLVVLGVVF-----GILIKRROQKIRKTYMRLLQETELVEPLTPSGAMPNOAQ 711
DB 640 TGIALGVPAGLIFELVLVFLGLMLVHRGLAIRKRAMRYLESGLSFEFLG-GEKTKVH 698
OY 712 MRILKETELRKVVLGSGAGFTYVKGWIPDGENVKIPVAIKVLRENTSFRANKEILDEA 771
DB 699 ARILKPEDLKRKIKPLGSGVFGVSKGFWIMEGETVKIPVAIKTIQDSSGRQTFTEIDHL 758
OY 772 YVAVGVSPIVSRLLGICLTSTVQLVTOAMPYGLLDHVRNRRGLSQDLNMCQIAK 831
DB 759 LMSGSLDHPYIVRLGICPGICQLVTOSSHGSLLEIRHKSLDPLQRLNMCVQIAK 818
OY 832 GMSYLEDVRLVHRDLAARVLYKSPNHVKITLFGIARLLDDETEYHADGKVPKMMAL 891
DB 819 GMYLLEHRVVKHKLARNLILKNDYQVQISDYGADLLYDDKIVYSEKTKIKMMAL 878
OY 892 ESTLRRFRFTHQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPORICTIV 951
DB 879 ESTLFRYTHQSDVWSYGVTVWEMWSFGAEFYASVQPOEVSVLEKEERLSQPAICTIDV 938
OY 952 YMIWVKCMMDISECRPRELSEFSMARPDPOREVTIQNDDLGAPAPLSTFTRSLLED 1011
DB 939 YMWVKCMMDENIRPFKEILASDFTMARDPRIYLVIRMEG-----ED 982
OY 1012 DMGDLVDAEEYLVPOQGFPCPPAPGAGVGHHRHSSSTRSGGDLTLGLPSEEEAP 1071
DB 983 SGNGEFL-----RGRSER---GILEADLEDEDEE-- 1008
OY 1072 RSLPLASSEGAGSDVFDGLMG---AAKGLSLPTHDPSPLO-----RYSEDPTV 1118
DB 1009 -----GLGDRFATPSLQPSPSMSTSPQINSYMWMTQLRYD----- 1044
OY 1119 PLPSETDGVVAPLTCSPQ- EYVNO-----PDVRRQPSREGPL--PAAR 1161
DB 1045 -FVSGCGHIGYLPMSFSPVDITIRQLMYORSRLSVRTLPRSAFRSSREAELECEDAQ 1103
OY 1162 PAGATLERAKTSLPGKNGVAVKDVAFGCAVENPEYLTPOGGAAPQPHRPAFSPAFONLY 1221
DB 1104 CAGIFPRV-----FGSERGN-----PQGG----- 1122
OY 1222 YMDQPEPERGAPSTFKGPTAENPE 1247
DB 1123 --QQRKLSTASSPSSFKTMADEDE 1146

```

```

OC Anopheles.
OX NCBI_taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4.
RA lycett G.O.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR HSP; A1301655; CAC35008.1; -.
DR HSP; P11362; JFGC.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; EGFR_L_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; Fu_7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ
SEQUENCE 1 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.8%; Score 2032.5; DB 5; Length 1433;
Best Local Similarity 32.6%; Pred. No. 2.8e-144;
Matches 471; Conservative 196; Mismatches 395; Indels 395; Gaps 36;

OY 26 CTGTDMKRLRPLASBETHLDMRLHLYOCQOVVQNLLETLYLPTNALSFLDIOIEVOGYVL 85
DB 1 CIGTNGMSVAPAREVHYKRLRDYNTCTYDGNLEITWIONITDLNFIQLIREVTGYL 60
OY 86 IAHNQVQVPLQRLIRYRGOLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140
DB 61 ISLYDLPOVILLPRQLIRGRTHFTKLNKEEAYGLFV-----SFSHMTL 104
OY 141 QLSLTELKGVLIQNPOLCYODITLMDQYIKASKF-IGITELNRSACHPCSPMC 199
DB 105 ELPLARDILGSGVFENNVLCHMKSIWBEIILAPQTSMOYTENFSSPERVCPCHPSC 104
OY 200 KGSRCWGESSEDCOSLTRVACAGCA--RCKGPLPTDCHEQCAAGCTGPKHSQCLCLH 257
DB 165 EVG-CWEGAHNCORFSLKNCSPQCSGCRKFGPRFRCCHLFCAGGCTGPTOSDCLCKN 223
OY 258 FNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCYTAAPNYLSTDVGCTLVCP 317
DB 224 FYDDGVCKQECRPMQIYNPTNYFMEBPNDKVAAGATCVKCP-EHLKMGACVVRKCPK 282
OY 318 HNGEVLAEDEGTQRCCKSKPCARVCGSLGEMHAREVAUYSANIQEAGCKKIFGSLAF 377
DB 283 GKMPQNSE-----CVPKGVCPKTCPEGIVH-----SDNGNKDCDTIILEGSLIIL 329
OY 378 PESFDGDPASNT-----APLPQEOQLVFTLEETIGLYLISAWPDSLPDLSVFONLQ 429
DB 330 DQSFDFGQVYTNFSGPRYIKIDPRLVLFSTYKELTGFINIQAHHPNFTTLVPRNLE 389
OY 430 VIRGILHNGAY-SLTLOGIGISWILGRSLREUGSLALIHNTHTLCEVHTVPMDOLEFRN 488
DB 390 VVGGRQLKENTLFAFSAVYIVKTSLSKLELSKRVNCSGIVILENSDLCFVEDIDMSEIKS 449
OY 489 PHQALLHTANREDECVEGEGLAGHOLCARGHMGCPPTQCVACOSOFIRQOECEVBEQVLO 548
DB 450 SDHEVAVQKNRATTECHBERGBCSEQSKAGCWGKPEQDLECKNVYKKGKCLDSCK--- 506

```

OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

```

QY 549 GLPREY-VNARHCLPCHPECOFONGSVTCFPEADOCVACAHYKDPFCVACRP----- 601
DB 507 SLPLYSVDSKTCGDCHQCKD-----FCYGNENNCSSCMVNDGRGCVACEPTTKAM 561
QY 602 -----SGVKPDLSTYMP1WKFPD----- 618
DB 562 NGTCINCHKTCVGRGPRDTIAPDGCISDKAIIISDAKIBERCLMKDESCPDGYSDYVL 621
QY 619 -EEG----- 621
DB 622 QEEGPLKQLSGKAVCRKCHPRCKCTGYGFHEQFCQECTGYKKGQCEDECPDYANEE 681
QY 622 --ACOPCPINCT-----HSCVDL-----D-----KCGPAEQ----- 646
DB 682 TRILPCHQEGRCHGLDHDHNECNLTFEODPRDNTTFCVANCASHPRYKFPDEA 741
QY 647 -----RASPLTSIVSAVGLLVLLVGLVPGI---LIRKROQKIRKXTM 687
DB 742 GKIGPYCSADSMQSLRIEPOTQVKIWMGSVALILLCVFEGIAFLPSRHNKDKDAVGM 801
QY 688 RRLQETELVEPLTPSGAMPNQAOMRLKETELRKVKYLSGAFCTVYKGIIMPDENVK 747
DB 802 TMLAGCEDSEPLRSNVGPNLTUKRIIKEAEIRRGVAGMGAFGRVFKGVMMPEGESVK 861
QY 748 IPVAIKVLRENTSPKANKELIDEAYVMAGVSPYVSRLIGICTSTVOLVTQIMPYGCLL 807
DB 862 IPVAIKVLWEMSGSSSKFLEBAIYMASVEHPNLLKLLAVCMTSOMMLIQTLMLGCLL 921
QY 808 DHVENRGRLCSODLLNWCQIAKMSYLEVDVRLVHRLDAARNVLYKSPNHVKTIDFGIA 867
DB 922 DYVENNKKDKIGSKALLNWSSTOJARQMAYLEERLLVHRDLAARNVLYQPSCVKIVFGIA 981
QY 868 RLIDIDEYHADGKVPKIMMALESLIRRFTHQSDVMSGVTVWELMTFCAKRYDGI 927
DB 982 KLBDSDSEYPAAGKMPKIMLALCIRHVFSTSKDVMAFGITIMELTYGARREYVNP 1041
QY 928 AREIPDLKEGRLPQPICTIDVYIMVCMIDSECRPRELIVSEFSRMAADPQRFV 987
DB 1042 AKDVPELEIGHKLPQPIDICSLDYVLCILSCVLDADARPTFKQALFAEKARDPGYL 1101
QY 988 VIONEDLGASPLDSTFYRSLIEDDMGDLV----- 1018
DB 1102 MI-----PGDKFMRPLSYTNDQEKDLIRTLAPVMAAAAAAAGASNVDPSTIA 1152
QY 1019 DAEEYLVPOQGFPCDPAPGAGMWHHRSSSTSSGGDDLTLGLEPSEBEAPRS----- 1073
DB 1153 ETDELQPKTRPSIMLPQPSA-----VEPS-DEMPKSLRYCK 1188
QY 1074 -PLAP---SEGASDVFDGDLGMAAGKLGSLPTHDPSPLOQRYSEDPVPLPSETDGYVA 1129
DB 1189 DPLKRDDETDGKGV-----GVGGR-----LNLPLDEDDVLM 1222
QY 1130 PLTSPQPEYVNPQVPRQPSPRGRLPAPARPAGATLERAKTSLPGKNGVVKVAFARG 1189
DB 1223 P-TCOSQ---NQS-----PPG---YMDLIGVPA 1243
QY 1190 AVENPEYD-----TPGGAAPQPPPPAPSPAFNLYVWDQDPPERGAPESTFKGT 1240
DB 1244 SVDNPEYLMGSTOAIAGLAQSGM--PHTPP-----PPTTPNGM 1280
QY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

```

RESULT 8
Q9UK79 PRELIMINARY: PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DB Heretatin.

```

```

GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF17761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FFC1BE347E2D030C CRC64.

```

```

Query Match 26.5%; Score 1802.5; DB 4; Length 419;
Best Local Similarity 86.3%; Pred. No. 1.1e-127;
Matches 340; Conservative 9; Mismatches 38; Indels 7; Gaps 2;

```

```

QY 1 MELALCRWGLLALLPPGAASVCTGTDMKRLPASPEHLDMLRHLYOGCOVVGNTL 60
DB 1 MELALCRWGLLALLPPGAASVCTGTDMKRLPASPEHLDMLRHLYOGCOVVGNTL 60
QY 61 ELTILPTASISFLDIDIEVGVYLIANOVQVPLQRLIVRGTLFEDNYALAVLNG 120
DB 61 ELTILPTASISFLDIDIEVGVYLIANOVQVPLQRLIVRGTLFEDNYALAVLNG 120
QY 121 DPLNNTPTVGTASPGKRELQRLSTELIKGVLQRPOLCYODTILMKDQYIANKSKF 180
DB 121 DPLNNTPTVGTASPGKRELQRLSTELIKGVLQRPOLCYODTILMKDQYIANKSKF 180
QY 181 IGITELNRSRACHPSPCKSRCKSGESSEDCQSLTRTVACAGCARCKGRLPTDCHQC 240
DB 181 LTLDITNSRACHPSPCKSRCKSGESSEDCQSLTRTVACAGCARCKGRLPTDCHQC 240
QY 241 AAGCTGPRHSCTCLAHNHSIGICELRCPALVTYNTDTFESMPNPEGRTTGASCVTACP 300
DB 241 AAGCTGPRHSCTCLAHNHSIGICELRCPALVTYNTDTFESMPNPEGRTTGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPPLNNOEVTABDGTORCEKSGPCARVCYGLGMEHLREVRAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPPLNNOEVTABDGTORCEKSGPCARVCYGLGMEHLREVRAVTSAN 360
QY 361 IQEPAGCKKIFGSLAFLEPESFGDPASNTAPLP 394
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLPLAP 387

```

```

RESULT 9
Q8R2X1 PRELIMINARY: PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
DB Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.

```

DR EMBL/ BC027080.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;  
Best Local Similarity 88.0%; Pred. No. 5,5e-123;  
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESIILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICT 948  
DB 1 MALESIILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICT 60  
QY 949 IDVMIMVKKMMIDSECRPRFRELVSERMAARDQRFVVIQNEDELGPSSPMDSTFYRSL 1008  
DB 61 IDVMIMVKKMMIDSECRPRFRELVSERMAARDQRFVVIQNEDELGPSSPMDSTFYRSL 120  
QY 1009 LEDDDMGDLVDAEETLVPOGFFCPDPAPGAGWVHHRRSSSTRSGGDDLTLGLEPSEE 1068  
DB 121 LEDDDMGDLVDAEETLVPOGFFCPDPAPGAGWVHHRRSSSTRSGGDDLTLGLEPSEE 180  
QY 1069 EAPRSPLAPSGAGSDVDFDGLGCAKAGLQSLPTHDSPLQRYSEDPVPLPSETDGYV 1128  
DB 181 EAPRSPLAPSGAGSDVDFDGLGCAKAGLQSLPTHDSPLQRYSEDPVPLPSETDGYV 240  
QY 1129 APLTCSPOPEYVNOQDVPQPSPREGBLPAARPAATLERAKTILSPKNGVYKDVFAFG 1188  
DB 241 APLTCSPOPEYVNOQDVPQPSPREGBLPAARPAATLERAKTILSPKNGVYKDVFAFG 300  
QY 1189 GAVENPEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPERGAAPSTFFKGTPTAENEY 1248  
DB 301 GAVENPEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPERGAAPSTFFKGTPTAENEY 360  
QY 1249 LGLDVPPV 1255  
DB 361 LGLDVPPV 367

RESULT 10  
086712 PRELIMINARY; PRT; 729 AA.  
ID 086712  
AC 086712;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Polypeptide.  
GN POLYPEPTIDE.  
OS Avian rous-associated virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.  
NC NCB1\_Taxid=11950;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94203659; Pubmed=8152791;  
RA Vennetrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
RA Johnson A., Beug H.;  
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
RT different transforming capacities.";  
RL Oncogene 9:1307-1320(1994).  
DR EMBL; S69372; AAC60725.1; --  
DR HSSP; P03322; 1A6S.  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR InterPro; IPR004028; Retro\_M.  
DR InterPro; IPR001245; Tyr\_Pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF02813; Retro\_M; 1.  
DR ProDom; PD000001; Euk\_Pkinase; 1.  
DR SMART; SMO0219; Tyrc; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 729 AA; 80649 MW; 84DF6914EFD1D63 CRC64;

Query Match 25.3%; Score 1720; DB 15; Length 729;  
Best Local Similarity 54.8%; Pred. No. 4.2e-121;  
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 569 PONGSVTFQGPBADCVAHAKDPFCVAPRPSVKRDLSTMPYWKPFDEGACQPCPI 628  
DB 11 PONTATPKTGP--DHCKMCAHFIDGPHCVKACPAVLGENDTL-VMKADANAVCOLCHP 197  
QY 629 NCTHSCVDLDKGCAPAEORASPLTISAVV-GILLVYVAVGVLGILKRQOKIRKTYM 687  
DB 198 NCTHSCVDLDKGCAPAEORASPLTISAVV-GILLVYVAVGVLGILKRQOKIRKTYM 253  
QY 688 RRLLOETELVPLTPSGAMPNOQMRILKETLRKYKVLGSGAFGVYKGIWIPGGENYK 747  
DB 254 RRLLOETELVPLTPSGAMPNOQMRILKETLRKYKVLGSGAFGVYKGIWIPGGENYK 313  
QY 748 IPVAIKVLRMTSPANKELIDEAVYVAVGSPYVSLGTLCTSTVOLVTLMPYGCIL 807  
DB 314 IPVAIKVLRMTSPANKELIDEAVYVAVGSPYVSLGTLCTSTVOLVTLMPYGCIL 373  
QY 808 DHVENRGRGLGSDLLNMCQIAGMSYLEDVRLVHRDLAARNVLYKSPNHVKITDPGLA 867  
DB 374 DHVENRGRGLGSDLLNMCQIAGMSYLEDVRLVHRDLAARNVLYKSPNHVKITDPGLA 433  
QY 868 RLDDIDETRYHADGAKVPIKMALESILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGI 927  
DB 434 RLDDIDETRYHADGAKVPIKMALESILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGI 493  
QY 928 AREIPDLLEKGERLPPPICTIDVYMTVKKMIDSECRPRFELVSERMAARDQRFV 987  
DB 494 AREIPDLLEKGERLPPPICTIDVYMTVKKMIDSECRPRFELVSERMAARDQRFV 553  
QY 988 VIQ-NEIDGPASPLDSTFYRSLLEDQDGLVDAEETLVPOGFFCPDPAPGAGMVMHR 1046  
DB 554 VIQ-NEIDGPASPLDSTFYRSLLEDQDGLVDAEETLVPOGFFCPDPAPGAGMVMHR 598  
QY 1047 HRSSSTRSGGDLTLGLEPSEEAAPRSP-----APSEGAGSDVDFDGLGCAKAGLQSL 1101  
DB 599 HRSSSTRSGGDLTLGLEPSEEAAPRSP-----APSEGAGSDVDFDGLGCAKAGLQSL 631  
QY 1102 PTHPSPLOQRYSEDPVPLPSET--DGVAAPLTCSPOPEYVNOQDVPQPSPREGBLPA 1159  
DB 632 PTHPSPLOQRYSEDPVPLPSET--DGVAAPLTCSPOPEYVNOQDVPQPSPREGBLPA 675  
QY 1160 ARPAGATLERAKTILSPKNGVYKDV-----AFGAVENPEYL 1197  
DB 676 ARPAGATLERAKTILSPKNGVYKDV-----AFGAVENPEYL 715

RESULT 11  
086714 PRELIMINARY; PRT; 567 AA.  
ID 086714  
AC 086714;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE V-erbB protein (Fragment).  
GN V-ERBB.  
OS Avian rous-associated virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.  
NC NCB1\_Taxid=11950;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94203659; Pubmed=8152791;  
RA Vennetrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
RA Johnson A., Beug H.;  
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
RT different transforming capacities.";  
RL Oncogene 9:1307-1320(1994).  
DR EMBL; S69372; AAC60727.1; --  
DR HSSP; P11362; IFGK.  
DR InterPro; IPR000719; Euk\_Pkinase.

DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00069; kinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR Tyrosine-protein kinase.  
 FT NON TER  
 SO SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;  
 Best Local Similarity 55.4%; Pred. No. 4e-121;

Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQCAAHYKDPFCVACPSGVKRDLSYMIWFPDEGACOPRINCTSCVDL 637  
 1 GP--DHCKCAHFIDGPHCVKACPGVIGENDTL-VMKYADANAACOLCHPCTRCKGKP 57  
 DB 638 DDKGCPAEQASPLTSIVSAVY-GILLVVLGVGILIKRQOKIRKRYTMRLLQETEL 696  
 58 GLEGCP---NGSKTSIAAGVVGILCLVVGILGILYRR-HYRKRTLRLLQEREL 113  
 QY 697 VEPLTPSGAMPQACMRILKETELRKVKVLSGAFGVYKGIWPDGENVKIPVAIKVL 756  
 DB 114 VEPLTPSGEAPQAHRLIKETEFKVKVLSGAFGVYKGLWPEGEKVIPIVAIKELR 173  
 QY 757 ENTSPKAKETLDEYVYMAAGVSPYVSRLLGICLTSTVOLYQMLPYGGLDHDVENEKR 816  
 DB 174 EATSPKAKETLDEYVYMAASVDNPRVCRLLGICLTSTVOLYQMLPYGGLDHDVENEKR 233  
 QY 817 LGSODLNMCMQIAKMSYLEDVRLVHRDLAARNVLYVSPNHVKTIDFGLALDIDETE 876  
 DB 234 IGSQVLNMCQIAKMSYLEDVRLVHRDLAARNVLYVTPGVKITTDFGLALDIDETE 293  
 QY 877 YHADGKPIKMALESILRRFTYQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLIE 936  
 DB 294 YHAEKGKPIKMALESILRRFTYQSDVSYGVTVWELMTFGSKPYDGIIPAREIPDLIE 353  
 QY 937 KGERLPQPICTIDVYIMVYKMMIDSECRPRFELVSEFSRMAPDRQRFVYIO-NEGL 995  
 DB 354 KGERLPQPICTIDVYIMVYKMMIDSECRPRFELVSEFSRMAPDRQRFVYIO-NEGL 995  
 QY 996 PASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFCCDDPAPAGAMVHRRHSSTSG 1055  
 DB 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSFT--- 454  
 QY 1056 GGDLLTGLPSEBEAPRSPPL-----APSEGAGSDVFDGDLGKAAGLQSLPTHDPSPLQ 1110  
 DB 455 -----SRTPLLSLSTSNNSATNCID-----RNGQGHVRDEDSFVQ 491  
 QY 1111 RYSEDPVPLPSET--DGVAAPLTCSPQPEYVNOQDVAPQPSPREGLPAPARAGATLE 1168  
 DB 492 RYSSDPVPLPSET--DGVAAPLTCSPQPEYVNOQDVAPQPSPREGLPAPARAGATLE 1168  
 QY 1169 RAKTLSPGKNGVYKDFV-----AFGAVENPEYL 1197  
 DB 527 ----TAWVQNGIYNNISLTALSKLPMSDRYQNSHSTADNPEYL 566

RESULT 12  
 Q8WYVO PRELIMINARY; PRT; 412 AA.  
 AC Q8WYVO;  
 DT 01-MAR-2002 (TREMBLrel. 20. Created)  
 DT 01-MAR-2002 (TREMBLrel. 20. Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21. Last annotation update)  
 DE Hypothetical 44.7 kDa protein.  
 GN PP3659.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
 RA Wan D.F., Gu J.R.;  
 RT "Novel human cDNA clones with function of inhibiting cancer cell  
 RT growth."  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF181349; AAL55856.1; -  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR00719; Euk\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR InterPro; IPR004019; YLP motif.  
 DR Pfam; PF00069; kinase; 1.  
 DR Pfam; PF02757; YLP; 2.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Hypothetical protein.  
 SO SEQUENCE 412 AA; 44702 MW; 034397FF3272D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;  
 Best Local Similarity 80.5%; Pred. No. 8.9e-120;  
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALSILRRFTYQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLIEKGERLPQPICT 948  
 DB 1 MALSILRRFTYQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLIEKGERLPQPICT 60  
 QY 949 IDVYIMVYKMMIDSECRPRFELVSEFSRMAPDRQRFVYIONEDLGPAASPLDSTFYRSL 1008  
 DB 61 IDVYIMVYKMMIDSECRPRFELVSEFSRMAPDRQRFVYIONEDLGPAASPLDSTFYRSL 120  
 QY 1009 LEDDDMGDLVDAEYLVPOQGFCCDDPAPAGAMVHRRHSSTSGGDLTLGLEPSEE 1068  
 DB 121 LEDDDMGDLVDAEYLVPOQGFCCDDPAPAGAMVHRRHSSTSGGDLTLGLEPSEE 180  
 QY 1069 EAPRSLPSPGAGSDVFDGDLGKAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGV 1128  
 DB 181 EAPRSLPSPGAGSDVFDGDLGKAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGV 240  
 QY 1129 APLTSPQPEYVNOQDVAPQPSPREGLPAPARAGATLERAKTLPCKNGVYKDFVAFG 1188  
 DB 241 APLTSPQPEYVNOQDVAPQPSPREGLPAPARAGATLERAKTLPCKNGVYKDFVAFG 300  
 QY 1189 GAVENPEYLTPOGGAAPQ-----HPPA---FSPAFDNL 1220  
 DB 301 GAVENPEYLTPOGGAALPTLLPSAOPSTTSITGRTYQSGGLPAPSKGHLRQRTOST 360  
 QY 1221 YYMD-QDPPER-----GAPSTFKGTPTAEN 1245  
 DB 361 WMTQCQPEQVRRSPDVSSGREGLTSGAKIKRMEGPTTSRGTCARN 410

RESULT 13  
 Q64895 PRELIMINARY; PRT; 962 AA.  
 AC Q64895;  
 DT 01-NOV-1996 (TREMBLrel. 01. Created)  
 DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21. Last annotation update)  
 DE Gag; V-erb-A; V-erb-B protein.  
 GN GAG; V-ERB-A; V-ERB-B.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 OC NCBI\_TaxID=11861;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90206603; PubMed=1969616;  
 RA Brunkin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;  
 RT "Six amino acids from the retroviral gene gag greatly enhance the  
 RT transforming potential of the oncogene V-erb-B.";

```
RL Oncogene 5:15-24(1990)
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CA36459.1; -
DR EMBL; X52211; CA36459.1; JOINED.
DR HSSP; P10828; NUL.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR000536; Hormone_rec_1ig.
DR InterPro; IPR001723; Slchrm_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_Csteroi.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR Prodom; PD000001; Euk_kinase; 1.
DR Prodom; PD000035; Znf_Csteroi; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; Tyrc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_KINASE_ATP; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS01009; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 6.6e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEEGRVQLPRE-YVNAR-HCLP-----CHPECO 568
DB 354 IEKQSEYLAFFHYIYRKNINPHFMSKLMKADLMICAYHARSRLHMKVECPPELS 413
QY 569 PONGSVTCFGEADQVACAHYKDPFCVACPSGVKPDLSYMBIMKFPDEGACQPCPI 628
DB 414 PQE-----VGP--DHCKMCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCOLCHP 465
QY 629 NCHSCVLDLDDKGPAPORASPLTISYSAV-GILLVVLGVVGLIKRQOKIRKXTM 687
DB 466 NCRGCGPGLGCP--NSKTPSIAAGVGLICLVVGLIGLYLRR-HIVRKTL 521
QY 688 RRLLOETELVEPLTPSGAMPNOAMRLIKETELRKVVLGSGAFGYKGIWDGENVK 747
DB 522 RRLLOEELVEPLTPSGEARPQAHRLIKETEFKVKVGLGAFGYKGLMPEGEKVT 581
QY 748 IPAIKVLRNTSEPKANKELIDEAYVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGCIL 807
DB 582 IPAIKELREATSPKANKELIDEAYVYVAGVSDNPHVCHLIGICLTSTVOLITQMLPYGCLL 641
QY 808 DHVRENGRGLSGQDLMLWMCQIAGMSYLEVRLVHRDLAARNVLYKSPNHVKTIDFGLA 867
DB 642 DYIREHNDNGSOYLMLWMCQIAGMNYLEERHMHVHDLAARNVLYVTPQHVKITDGLA 701
QY 868 RLIDIDETEHADGKVPKIMMALESILRRFTHQSDVMSYGVTVMLMFGAPYDGP 927
DB 702 KQCGADEKEVHAEGKVPKIMMALESILHRIYTHQSDVMSYGVTVMLMFGSPYDGP 761
QY 928 AREIPDLLEKGERLPQPPICITIDVYIMVYKCMWIDSECRPRFELVSEFSMARDPQRFV 987
DB 762 ASEISSVLEKGERLPQPPICITIDVYIMVYKCMWIDSECRPRFELVSEFSMARDPQRFV 821
QY 988 VIO-NEDLGASPLDSTFFVYSLLEDDMDGLVDAEETLVPOQGFCDPAPAGAGMYHR 1046
DB 822 VIOGDEKMLPSPTDSKFYRLLMEEDMEIDVDAEYLVPHQGF----- 866
QY 1047 HRSSSTRSGGDLTLGLPESEEARPLAPSEGAGSDVFDGDMGAAKGLQSLPTHP 1106
DB 867 -NSPST-----SRTPLLSSLSATSN-----NSATYCIDRNGH-- 898
```

```
QY 1107 SPLORSEDPVPLPSETDGVAPLTCSPQPEYVQNPVRPQPSREGPLPAARPAQAT 1166
DB 899 -----PVREDGFL-----PAPEYVQ--LMPKKESTANVQNIYVISLT 936
QY 1167 -LERAKTLSPGKGVKQVAFAPGAVENPEYL 1197
DB 937 AISKLPMDSRVQ-----SHSTAVNDPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TREMBLrel. 01. Created)
DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20. Last annotation update)
DE Avian erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
VIRUSES; Retroid viruses; Retroviridae; Avian type C retroviruses.
OC NCI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CA330024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR Prodom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS01009; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferrase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DC8C8CA078AF4 CRC64;

Query Match 24.2%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.2e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADQVACAHYKDPFCVACPSGVKPDLSYMBIMKFPDEGACQPCPINCNSVDL 637
DB 1 GP--DHCKMCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCOLCHPCTRGCKP 57
QY 638 DDKCPAQRASPLTISYSAV-GILLVVLGVVGLIKRQOKIRKXTMRRLLOETEL 696
DB 58 GLEGCP--NSKTPSIAAGVGLICLVVGLIGLYLRR-HIVRKTLRRLLOEREL 113
QY 697 VEPLTPSGAMPNOAMRLIKETELRKVVLGSGAFGYKGIWDGENVKIPVAIKYLR 756
DB 114 VEPLTPSGEARPQAHRLIKETEFKVKVGLGAFGYKGLMPEGEKVTIPVAIKELR 173
QY 757 ENTSPKANKELIDEAYVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGCILDHVRENGR 816
DB 174 EATSPKANKELIDEAYVYVAGVSDNPHVCHLIGICLTSTVOLITQMLPYGCLLDYIREHND 233
QY 817 LGSODLMLWMCQIAGMSYLEVRLVHRDLAARNVLYKSPNHVKTIDFGRLIDIDETE 876
DB 234 IGSQYLLMNCVQIAGMNYLEERHMHVHDLAARNVLYKTPQDVKITIDGLKQCGADEKE 293
QY 877 YHADGKVPKIMMALESILRRFTHQSDVMSYGVTVMLMFGAPYDGPAPAREIPDLLE 936
DB 294 YHAEKGVPIKIMMALESILHRIYTHQSDVMSYGVTVMLMFGSPYDGPAPASEISSVLE 353
QY 937 KGERLPQPPICITIDVYIMVYKCMWIDSECRPRFELVSEFSMARDPQRFVIO-NEDLG 995
DB 354 KGERLPQPPICITIDVYIMVYKCMWIDSECRPRFELVSEFSKMARDPQRFVIOGDERMH 413
```

```

QY 996 PASPLDSTFYRSLLIEDDDMDGLVDAEEXLVPOQGFCDPPAPAGAGMHHRRSSSTRSG 1055
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 414 LPSFTDSEFYRTLMEEBEMEDIYDAEDYLVEHQGF-----NGST----- 454
QY 1056 GGDLLTLGLEPSEEBEAPRSPL-----APSEGAGSDVFDGDLGKAAGLQSLPTHDPSPLQ 1110
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 455 -----SRTPLLSLSATSNNSTXNCIDRNGC-----H----- 481
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNOVDVRRPQPSREGPLPAARAGAT-LER 1169
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 482 -----PVREDEGFL-----PAPXYVNO-LMPKKPSTAMVOIQIYNYISLTAISK 523
QY 1170 AKTLSPGKGVKVDVFAFGAVENPEYL 1197
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 524 LPMDSRYGN-----SHSTAVDNEPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
CN EGF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibhe N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/10J, 129/SVJ, AND 129/SVEVYAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Konno H., Aachi J., Fukuda S., Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana I., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Flaischmann W., Gaasterland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J., Schmitt L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G., Blake J., Boffelli P., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein W.J., Bult C., Fletcher C., Fujita M., Gaitanaris M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontseki S., Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DB EMBL; AF124513; AAD4149.1; -.
DB EMBL; AF275366; AAG28047.1; -.
DB EMBL; AF275364; AAG28047.1; JOINED.
DB EMBL; AF275365; AAG28047.1; JOINED.
DB EMBL; AK004944; BAB33688.1; -.
DB EMBL; AK004883; BAB33684.1; -.
DB EMBL; AK004911; BAB23662.1; -.
DB MGI; MGI:95294; Egtf.
DR InterPro; IPR00494; EGF_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.4%; Score 1524.5; DB 11; Length 655;
Best Local Similarity 44.1%; Pred. No. 2,1e-106;
Matches 280; Conservative 100; Mismatches 240; Indels 15; Gaps 5;

QY 11 LLALLPPGAA--STOVCTGDMKLRLPASPEYTHLMDRLHYOGCOVVOGNLELYLPTN 68
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 14 LLTALCAAGALAEKXKVCQGSNRLTQGTGFEDHFLSLQRMVYNNCEVVLNLEITYVQRN 73
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 69 ASLSFLQDIQEVQGVYLIANQVROVPLQRLRIYVGTQLFENYVLAIVDNDPLNNTTP 128
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 74 YLSFLKTIQEVAGVGLTALNTVERIPLNQIIRGNALYENTVALAILSN----- 124
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 129 VTGASPGRLRELQSLREILKGVLIQRNPOLCVQDTLWMDQYIKANSKFIQITELNR 188
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 125 -YGNKRTLRLEIPNMLQELILGAVFSNNPLCNMDITIQMDIYQVNFMSNMSDLOGH 183
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 189 SRACPCSPMKGSKRCSWSESSBDCSLTRTVACGCA-RCKGPLPTDCCHECCAGCTGP 247
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 184 PSSCPKCPSCPCNGSCWGGEGENCKLTKIICAQGCSHRCRGRSPDCCNCCAGCTGP 243
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 248 KHSDDLALHFRHSGICELHCPALVTYNTDFESHPNREGRTFGASCTYACPYNYLSD 307
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 244 RESDPLVCQKQDEATCKDTCPPLMLYNPTTYQMDVNPBGKYSFGATCYKCKPRNVYVD 303
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 308 VGSCTLVCPPLHNQVETADGTORCEKSKPCARVCYGLGMEHLREVRATVSANIOEFAC 367
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 304 HGSCYRACGPDIYEV-EEDGIRKCKCKGCPCKKVCNGIGIGFKOTLSINAINIKFKYC 362
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 368 KKIIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBITGYLYISAMPDSLPLDSVFQ 427
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 363 TALSDDLHLPLVAFKGSFTTRPPLDPRELEIKTVKEITGFLLIQAMPDNTDLHAFEN 422
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 428 LQVIRGRILHNGAYSILTQIGISVWLGRLSRLSELGSLALIHNNTHLCFVHTVPMDQLR 487
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 423 LEIRIGRTRKHQGFSLAVVGLNITSILGLRSLKEISDGVIIISGNRNLCYANTINMKKFLG 482
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 488 NPHQALHTARPREBECVGEGLAQHQLCARGHGCMWGPPTQCNCSQPLRGQCEVEECRYL 547
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 483 TPNQTKTKMNAEADKCAVNHVCCNPCLSSBCCWCPPEPDQVSCCNVSRGRCEVECNIL 542
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 548 QGLPREYVNAHRLCPHEPCQPNQSVTCFGEADQCYACAHYKDPPEFCVAPCPSPGVKPD 607
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 543 EGEPEPEFVNSBECICHECECLPQANINICTGNGPNCIGQCAHYIDGPHCVKCPAGIMGE 602
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 608 LSYMPIKMFPEBEGACQCPPINCHSCYDLDDKGC 642
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 603 NNTL-VMKYADANNVCHLCHANCTYGACAGPGLGC 636
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

Search completed: July 22, 2003, 09:00:28  
 Job time : 53.3575 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds

(without alignments)  
4403.399 Million cell updates/sec

Title: SEQ4-149-163-12

Perfect score: 6810  
Sequence: 1 MELALCRWGLLLALPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6740	99.0	1255	21	AAV92620 Human heregulin 2
2	6740	99.0	1255	22	AAE12130 Human tyrosine kin
3	6740	99.0	1255	22	AAE60167 HER2 transgene pla
4	6740	99.0	1255	23	AAU74545 Human HER2 (ErbB2)
5	6734	98.9	1255	17	AAW01111 HER-2/neu protein.
6	6734	98.9	1255	20	AAW92406 Human HER-2/neu pr
7	6734	98.9	1255	21	AAAB21198 Amino acid sequenc
8	6734	98.9	1255	21	AAAB84780 Human HER-2/neu pr
9	6734	98.9	1255	22	AAAB8458 HER2/neu amino aci
10	6734	98.9	1255	22	AAAG8267

11	6734	98.9	1255	23	AAE24067 Human Her-2 protei
12	6734	98.9	1255	23	AAE20479 Human Her-2/neu pr
13	6734	98.9	1255	23	AAW51143 Human Her-2/neu pr
14	6734	98.9	1255	23	AAU77114 Human Her-2/neu po
15	6691	98.3	1433	14	AAE39568 Sequence of c-erbB
16	6570	96.5	1223	23	AAU98923 Human breast cance
17	6417	94.2	1200	21	AAAB21208 Human HER-2/neu pr
18	5945.5	87.3	1256	23	AAAB21199 Rat HER-2/neu prot
19	5945.5	87.3	1256	23	AAW51144 Mouse Her-2/neu onc
20	5925.5	87.0	1256	22	AAAB21206 Amino acid sequenc
21	5925.5	87.0	1256	22	AAAG2860 Mouse Her-2/neu pr
22	5925.5	87.0	1256	23	AAW51151 Mouse Her-2/neu on
23	4820	70.8	919	23	AAAB21203 Human HER-2/neu fu
24	4820	70.8	919	23	AAW51148 Her-2/neu extracel
25	4075.5	59.8	920	23	AAW51152 Mouse Her-2/neu ex
26	4075.5	59.8	926	23	AAW51153 Mouse Her-2/neu ex
27	3704	54.4	712	21	AAAB21204 Human HER-2/neu fu
28	3704	54.4	712	21	AAW51149 Her-2/neu extracel
29	3558	52.2	782	18	AAW19764 Her2-GW-CSF immuno
30	3556	52.2	653	21	AAAB21200 Extracellular HER-
31	3556	52.2	653	23	AAW51145 Human ErbB2 oncopr
32	3518	51.7	645	22	AAAB61593 Human ErbB2 extrac
33	3453	50.7	951	21	AAV44993 DC8CFV-erbB2EC fu
34	3350	49.2	624	11	AAE08222 Amino acid sequenc
35	3160	46.4	1210	21	AAAB19259 Human EGF receptor
36	3160	46.4	1210	21	AAE23019 Human Her-1 protei
37	3160	46.4	1210	23	AAW50768 Human epidermal gr
38	3160	46.4	1210	22	AAAB6420 Amino acid sequenc
39	3158	45.8	1210	22	ABP51768 Human epidermal gr
40	3119	45.8	583	23	AAE20483 Human protein for
41	3084	45.3	587	23	AAE20481 Human protein for
42	3084	45.3	589	23	AAE20484 Human protein for
43	3083	45.3	600	23	AAE20482 Human protein for
44	3083	45.3	600	23	AAE20482
45	3083	45.3	600	23	AAE20482

#### ALIGNMENTS

RESULT 1	AAV92620	standard; Protein; 1255 AA.
ID	AAV92620	
XX	AAV92620;	
AC	10-AUG-2000	(first entry)
XX		
DT		
XX		
XX		
DE		
XX		
XX		
KW	Her2; vaccination; cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
XX		
OS	Homo sapiens.	
XX		
XX		
XX		
FT	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= "N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT /label= Cysteine\_rich\_domain  
 FT 210..224  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 250..264  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 324..483  
 FT /label= Ligand\_binding\_domain  
 FT 325..339  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 369..383  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 465..479  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 484..623  
 FT /label= Cysteine\_rich\_domain  
 FT 579..593  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 624..654  
 FT /label= Transmembrane\_domain  
 FT 632..652  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 653..667  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 655..1010  
 FT /label= Tyrosine\_kinase\_domain  
 FT 661..675  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 695..709  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 710..730  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 1011..1235  
 FT /label= C-terminal\_domain  
 FT  
 PN WO200020027-A2.  
 PD 13-APR-2000.  
 PF 05-OCT-1999; 99WO-DK00525.  
 PR 05-OCT-1998; 98DK-0001261.  
 PR 20-OCT-1998; 98US-0105011.  
 XX (MEBI-) M & E BIOTECH AS.  
 PA Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,  
 PI Gaum A, Birk P, Karlsson G;  
 PI WPI: 2000-349917/30.  
 DR N-PSDB; AAA09455.  
 DR  
 XX Inducing immune responses to weakly immunogenic, tumor associated  
 PT peptide antigens for the treatment of breast and prostate cancer  
 PT  
 PS Claim 62; Page 193-198; 220pp; English.  
 XX  
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL  
 CC response. Subdominant CTL epitopes, antibody binding regions and  
 CC cysteine residues involved in disulfide bonds are preserved in the  
 CC immunogenized forms. Regions suitable for the insertion of foreign T  
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
 CC cell-associated peptide antigens (PA) such as those associated with  
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
 CC The method comprises effecting simultaneous presentation by antigen  
 CC producing cells (APCs) of the animals immune system of: (1) at least 1  
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
 CC B-cell group derived from the cell-associated PA; and (2) at least 1  
 CC first T helper cell group which is foreign to the animal. Analogues of  
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
 CC part of all known and predicted CTL and B-cell epitopes of the respective  
 CC PA and including at least one foreign T helper epitope are also claimed.  
 CC The method is used to treat prostate, prostate/preast or breast cancer  
 CC when the PA is human PSM, FGF8b and Her2, respectively.  
 CC XX

SQ Sequence 1255 AA:

Query Match 99.0%; Score 6740; DB 21; Length 1255;  
 Best Local Similarity 99.0%; Pred. No. 0;  
 Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASVCTGTDMKRLPASSETHLDMRLHYOGGVQGNL 60  
 DB 1 MELAALCRWGLLALLPPGAASVCTGTDMKRLPASSETHLDMRLHYOGGVQGNL 60  
 QY 61 ELTYLPTNASLFLDIOEVGVYLIANQVQVPLQRLRIYRGTOLPEDNYALVDNG 120  
 DB 61 ELTYLPTNASLFLDIOEVGVYLIANQVQVPLQRLRIYRGTOLPEDNYALVDNG 120  
 QY 121 DPLNNTPTVTGASPGGLRELQLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180  
 DB 121 DPLNNTPTVTGASPGGLRELQLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180  
 QY 122 DPLNNTPTVTGASPGGLRELQLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180  
 DB 122 DPLNNTPTVTGASPGGLRELQLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180  
 QY 181 IGTIELNRRACHPCSPMKCKGRCESEDDCSLTTRYCAGGACARCKPLPTDCCHEQC 240  
 DB 181 IGTIELNRRACHPCSPMKCKGRCESEDDCSLTTRYCAGGACARCKPLPTDCCHEQC 240  
 QY 181 IGTIELNRRACHPCSPMKCKGRCESEDDCSLTTRYCAGGACARCKPLPTDCCHEQC 240  
 DB 181 IGTIELNRRACHPCSPMKCKGRCESEDDCSLTTRYCAGGACARCKPLPTDCCHEQC 240  
 QY 241 AAGCTGPKASDCLACIHPNHSIGICELHCALVTYNTDESNMPNPGRTTFGASCTYACP 300  
 DB 241 AAGCTGPKASDCLACIHPNHSIGICELHCALVTYNTDESNMPNPGRTTFGASCTYACP 300  
 QY 301 YNYLSTDVSGCTLVCEPLHNOEYTABDGTORCEKSPCARVCYGLGMEHLREVAATYSAN 360  
 DB 301 YNYLSTDVSGCTLVCEPLHNOEYTABDGTORCEKSPCARVCYGLGMEHLREVAATYSAN 360  
 QY 361 IOEPFACCKKIFGSLAFLPESFDGDPASNTAPLQPELOVETLEETGYLYISAWPDSL 420  
 DB 361 IOEPFACCKKIFGSLAFLPESFDGDPASNTAPLQPELOVETLEETGYLYISAWPDSL 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLISWLGRLSELGSLALIHNNTHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLISWLGRLSELGSLALIHNNTHLCFVHTV 480  
 QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHGKMGPTQVCNCSQFLRGQEC 540  
 DB 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHGKMGPTQVCNCSQFLRGQEC 540  
 QY 541 VEECKRYLQCLPREYVNAARHCLPCHPECQONSVTTCFGBADQCAVCAHYKDPFCVANC 600  
 DB 541 VEECKRYLQCLPREYVNAARHCLPCHPECQONSVTTCFGBADQCAVCAHYKDPFCVANC 600  
 QY 601 PSGVPRDLSYMPIMWFPDEEGACOPCPINCHTSCVDLDDKGPABGRASPLTSYAVVG 660  
 DB 601 PSGVPRDLSYMPIMWFPDEEGACOPCPINCHTSCVDLDDKGPABGRASPLTSYAVVG 660  
 QY 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVBLTPSGAMPNQAKRLKETEL 720  
 DB 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVBLTPSGAMPNQAKRLKETEL 720  
 QY 721 RRVKVLGSGAFETVYKGIWIPDGENVKIPVALIKVIRENTSPANKREILDEAVVMGVGSP 780  
 DB 721 RRVKVLGSGAFETVYKGIWIPDGENVKIPVALIKVIRENTSPANKREILDEAVVMGVGSP 780



```
QY 781 YSRLLGICLTSTVOLVLTQMPYGCCLLDHRENRGRIGSODLNMWQIAKMSYLEDDR 840
DB 781 YSRLLGICLTSTVOLVLTQMPYGCCLLDHRENRGRIGSODLNMWQIAKMSYLEDDR 840
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLDDIDETEXHAOGGKPIKMMALLESILRRFT 900
DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLDDIDETEXHAOGGKPIKMMALLESILRRFT 900
QY 901 HQSDVMSYGVTVLWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVNCMM 960
DB 901 HQSDVMSYGVTVLWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVNCMM 960
QY 961 IDSECPREFRELVSERFARMARDQRFVYIIONEDIGRPSPLDSTRYRSLLEDDMDGLVDA 1020
DB 961 IDSECPREFRELVSERFARMARDQRFVYIIONEDIGRPSPLDSTRYRSLLEDDMDGLVDA 1020
QY 1021 EEYLVPOQGFCCPPAPGAGAMVHRRSSSTRSGGDLTLGLERSEEARPRSLARSEG 1080
DB 1021 EEYLVPOQGFCCPPAPGAGAMVHRRSSSTRSGGDLTLGLERSEEARPRSLARSEG 1080
QY 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQDVPVPPSPSPREBGLPAPARPGATIERAKTISPGKKGVYKDYFARCGAVENPEYLTPO 1200
DB 1141 NQDVPVPPSPSPREBGLPAPARPGATIERAKTISPGKKGVYKDYFARCGAVENPEYLTPO 1200
QY 1201 GGAAPPHPPAPSPAFDNLYYWDQDPPERGAPESTKGTPTAENPEYLGADV 1255
DB 1201 GGAAPPHPPAPSPAFDNLYYWDQDPPERGAPESTKGTPTAENPEYLGADV 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
AC AAE12130;
DT 18-DEC-2001 (first entry)
DE Human tyrosine kinase-type receptor, HER-2.
XX
XX Therapeutic compound; major histocompatibility complex; vaccine;
XX antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
XX adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
XX antigen presenting cell; human; tyrosine kinase-type receptor.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Region 774..782
XX FT /note="Antigenic epitope"
XX PN MO200168677-A2.
XX PD 20-SEP-2001.
XX PF 16-MAR-2001; 2001WO-US40328.
XX PR 16-MAR-2000; 2000US-0527487.
XX PA (GEN2 ) GENZYME CORP.
XX PI Nicolette CA;
XX DR MPI: 2001-616284/71.
XX DR N-PSDB; AADI9731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
XX for use in adoptive immunotherapy, has enhanced binding to major
XX histocompatibility molecules and enhanced immunoregulatory properties
XX
```

```
XX
XX Claim 4; Page 63-67; 69pp; English.
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 99.0%; Score 6740; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
QY 1 MEIAALCRMGLLALLPFGAASSTQVCTGTMKRLPASPEHMLRLHYGCGVQGNL 60
DB 1 MEIAALCRMGLLALLPFGAASSTQVCTGTMKRLPASPEHMLRLHYGCGVQGNL 60
QY 61 ELTYLPINASLSFLQDIQEVQGVYLIANQVQVPLQRLIRVGTQLEFDNYALAVLDNG 120
DB 61 ELTYLPINASLSFLQDIQEVQGVYLIANQVQVPLQRLIRVGTQLEFDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGIRELQLSLTETILKGVYLIQNNPOLCYODTLIMXDI FHKNO LA 180
DB 121 DPLNNTPTVYGASPGGIRELQLSLTETILKGVYLIQNNPOLCYODTLIMXDI FHKNO LA 180
QY 181 IGITELNRSRACHPCSPMCKSGSRCKGSESSDCSLTRTVACGACRCKGPLPTCCHEOC 240
DB 181 LTLIDTRNSRACHPCSPMCKSGSRCKGSESSDCSLTRTVACGACRCKGPLPTCCHEOC 240
QY 241 AACCTGPKHSDDLACLHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AACCTGPKHSDDLACLHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNLTSTDVGSCTIVCPAHNDEVTAEDTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNLTSTDVGSCTIVCPAHNDEVTAEDTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
QY 361 IOEFAGCKTIFGSLAFLPESFDGDPASNTAPLOPEQLQVEETLEITGYLYISAMPDLP 420
DB 361 IOEFAGCKTIFGSLAFLPESFDGDPASNTAPLOPEQLQVEETLEITGYLYISAMPDLP 420
QY 421 DLSVFQNLQVIRRIILHNGAVSLTLQGLGISWLGSLRLRELSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRRIILHNGAVSLTLQGLGISWLGSLRLRELSGLALIHNNTHLCFVHTV 480
QY 481 PMDQLFRNPHQALLHFRANRPEDECVEGLACHOLCARHGCMGPGPTCCVNCQFLRQEC 540
DB 481 PMDQLFRNPHQALLHFRANRPEDECVEGLACHOLCARHGCMGPGPTCCVNCQFLRQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPEQOPNGSVTCGPAPDQVCAAHKDPFCVARC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHPEQOPNGSVTCGPAPDQVCAAHKDPFCVARC 600
QY 601 PSGVKPDLSTYMPIMKFPDEBGACQPCPINTHSCVDLDKGCAPAGRASPLTISVAVNG 660
DB 601 PSGVKPDLSTYMPIMKFPDEBGACQPCPINTHSCVDLDKGCAPAGRASPLTISVAVNG 660
QY 661 ILLVVLGVVFGILIRROOKIRKTYMRRLQETELVEPLTPSGAMPNQAOMRILKXETEL 720
DB 661 ILLVVLGVVFGILIRROOKIRKTYMRRLQETELVEPLTPSGAMPNQAOMRILKXETEL 720
```

```

Qy 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
    |||||
Db 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
Qy 781 YVSRLLGICLTSTVQVLTQVLTQMPYCCLDHRENRGRIGSODLLWMCQIAKMSYLEDVR 840
    |||||
Db 781 YVSRLLGICLTSTVQVLTQVLTQMPYCCLDHRENRGRIGSODLLWMCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLVKSPPNHVKITDGLARLLDIDETEYHADGKVPIMMALESTLRRFT 900
    |||||
Db 841 LVHRDLAARNVLVKSPPNHVKITDGLARLLDIDETEYHADGKVPIMMALESTLRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVXCM 960
    |||||
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVXCM 960
Qy 961 IDSECRFRFRELVESEFRMARDPQRFVYIIONEDLGFPASPLDSTFYRSLLDDMDGLVDA 1020
    |||||
Db 961 IDSECRFRFRELVESEFRMARDPQRFVYIIONEDLGFPASPLDSTFYRSLLDDMDGLVDA 1020
Qy 1021 EBYLVPOQGFPCPPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEEARSPPLASEG 1080
    |||||
Db 1021 EBYLVPOQGFPCPPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEEARSPPLASEG 1080
Qy 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPVPLPSETDGVVAPLTCSPOBEYV 1140
    |||||
Db 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPVPLPSETDGVVAPLTCSPOBEYV 1140
Qy 1141 NQPPVRQPPSPRPGPLPAARPAATLEBRATLSPGKGVYKQVPAFAGAVENPEYLTPO 1200
    |||||
Db 1141 NQPPVRQPPSPRPGPLPAARPAATLEBRATLSPGKGVYKQVPAFAGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAFSPAFDNLVYWDOPPERGAPSTFGKPTJANPEYLGADV 1255
    |||||
Db 1201 GGAAPQHPPPAFSPAFDNLVYWDOPPERGAPSTFGKPTJANPEYLGADV 1255

RESULT 3
AAB60167 standard; Protein: 1255 AA.
AC AAB60167;
DT 03-APR-2001 (first entry)
DE HBR2 transgene plasmid construct encoded protein.
KW Human; HBR2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
    antibody.
OS Homo sapiens.
OS Synthetic.
PN MO200100244-A2.
XX 04-JAN-2001.
XX 23-JUN-2000; 2000WO-US17229.
XX 25-JUN-1999; 99US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX (GETH ) GENENTECH INC.
XX Ericsson S, Schwall R;
XX WPI, 2001-061962/07.
XX N-PSDB; AAF24297.
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

```

```

PS Example 3; Fig 4; 92pp; English.
XX The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HBR2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 99.0%; Score 6740; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
Qy 1 MELAALCRKGLIALLPPGAASVCTCTDMLRLPASPEHMLMLRLYGCQOVQGNL 60
    |||||
Db 1 MELAALCRKGLIALLPPGAASVCTCTDMLRLPASPEHMLMLRLYGCQOVQGNL 60
Qy 61 ELTYLPTNASLSFLODIOEVQGYVLIANQVQVPLQRLRIVRGSTOLFEDNYALAVLDNG 120
    |||||
Db 61 ELTYLPTNASLSFLODIOEVQGYVLIANQVQVPLQRLRIVRGSTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVYGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
    |||||
Db 121 DPLNNTTPVYGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
Qy 181 IGITELRSRACHCSFMCCKSRGWSSSEPCOSLTITVCAGGCARCKGRLPTCCHEOC 240
    |||||
Db 181 IGITELRSRACHCSFMCCKSRGWSSSEPCOSLTITVCAGGCARCKGRLPTCCHEOC 240
Qy 241 AAGCTGKHSDDLACLHFNHSGICELHCPALVTNTTFESMPNPEGRTYGASCVTRCP 300
    |||||
Db 241 AAGCTGKHSDDLACLHFNHSGICELHCPALVTNTTFESMPNPEGRTYGASCVTRCP 300
Qy 301 YNYLSTDVGSCTVLCPLHNOEVTAEDETQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
    |||||
Db 301 YNYLSTDVGSCTVLCPLHNOEVTAEDETQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEOQLQVFTELEITGYLIYSAMPDLP 420
    |||||
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEOQLQVFTELEITGYLIYSAMPDLP 420
Qy 421 DLSVFQNLQVYRGRILHNGAYSLTQIGISWLGRLSRLRELSGSLALIHNTHLCFYHTV 480
    |||||
Db 421 DLSVFQNLQVYRGRILHNGAYSLTQIGISWLGRLSRLRELSGSLALIHNTHLCFYHTV 480
Qy 481 PMDQLFNNPQALLHTANRPEDECVGGLACHQICARGHCWGPPTQCVNCSQFLRGQEC 540
    |||||
Db 481 PMDQLFNNPQALLHTANRPEDECVGGLACHQICARGHCWGPPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLIQGLPREYVNAARHCLPCHPECOPONGSYTCGPEPADOCVACAHYDPPFCVARC 600
    |||||
Db 541 VEECRVLIQGLPREYVNAARHCLPCHPECOPONGSYTCGPEPADOCVACAHYDPPFCVARC 600
Qy 601 PSGVYPLSTYMPIKFPEDEGACOPCPINCHSCVVDLDKGCAPAEORASPLTISVSAVVG 660
    |||||
Db 601 PSGVYPLSTYMPIKFPEDEGACOPCPINCHSCVVDLDKGCAPAEORASPLTISVSAVVG 660
Qy 661 ILVVVVLGVVFGILIKRQOKIRKRYTNRRLQETELVEPLTPSGAMPNOAQMRLKETEL 720
    |||||
Db 661 ILVVVVLGVVFGILIKRQOKIRKRYTNRRLQETELVEPLTPSGAMPNOAQMRLKETEL 720
Qy 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
    |||||
Db 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
Qy 781 YVSRLLGICLTSTVQVLTQVLTQMPYCCLDHRENRGRIGSODLLWMCQIAKMSYLEDVR 840
    |||||
Db 781 YVSRLLGICLTSTVQVLTQVLTQMPYCCLDHRENRGRIGSODLLWMCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLVKSPPNHVKITDGLARLLDIDETEYHADGKVPIMMALESTLRRFT 900
    |||||

```

Db 841 LVHDDLAARNVLSVSNHVKITTDGLARLLIDETETHTADGKVIKMMALSLIRRF 900  
 Qy 901 HQSDWVSQVTVWEIMTEGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKKCM 960  
 Db 901 HQSDWVSQVTVWEIMTEGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKKCM 960  
 Qy 961 IDSECRPRFRLVSEFSMADPQRFVVIQNEIDLCPASPLDSTFRRSLLEDDMDLVDA 1020  
 Db 961 IDSECRPRFRLVSEFSMADPQRFVVIQNEIDLCPASPLDSTFRRSLLEDDMDLVDA 1020  
 Qy 1021 EELVPOQGFCCPDPAAGGVMHHRSSSTRSGGGLTTLGLESEBEAPSPPLASEG 1080  
 Db 1021 EELVPOQGFCCPDPAAGGVMHHRSSSTRSGGGLTTLGLESEBEAPSPPLASEG 1080  
 Qy 1081 AGSDVFDGDLQMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
 Db 1081 AGSDVFDGDLQMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
 Qy 1141 NQDPVRPQPSPRREGPLPAARPAAGTLERAKTSLGKXGVVDVAFGAVENPEYLTPO 1200  
 Db 1141 NQDPVRPQPSPRREGPLPAARPAAGTLERAKTSLGKXGVVDVAFGAVENPEYLTPO 1200  
 Qy 1201 GGAAPQHPHPPAFSPAFDNLVYMDPPERGAPSTFKGTPAENPEYLGIDVPV 1255  
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYMDPPERGAPSTFKGTPAENPEYLGIDVPV 1255  
 RESULT 4  
 AAU74545 ID AAU74545 standard; Protein, 1255 AA.  
 AC AAU74545;  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human HER2 (ErbB2) polypeptide.  
 XX  
 KM Human; HER2; ErbB; epidermal growth factor receptor; receptor;  
 KM anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KM stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KM thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KM glioma disorder; astrocytoma disorder; hypothalamic disorder;  
 KM glandular disorder; macropapillary disorder; epithelial disorder;  
 KM stromal disorder; blastocoele disorder; inflammatory disorder;  
 KM angiogenic disorder; immunological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002001587-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 16-MAR-2001; 2001US-0811123.  
 XX  
 PR 16-MAR-2000; 2000US-189844P.  
 PR 05-OCT-2000; 2000US-238327P.  
 XX  
 PA (ERIC/) ERICKSON S.  
 PA (SCHW/) SCHWALL R.  
 PA (SLIW/) SLIWOWSKI M.  
 XX  
 PI Erickson S, Schwall R, Sliwowski M;  
 XX  
 DR WPI; 2002-163686/21.  
 DR N-PSDB; ABK14058.  
 XX  
 PT Treating tumour characterised by overexpression of epidermal growth  
 PT factor receptor. ErbB and cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal  
 XX  
 PS Example 3; Fig 7; 93pp; English.  
 CC The invention relates to creating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytic, hypothalamic, glandular, macropapillary,  
 CC epithelial, stromal, blastocoele, inflammatory, angiotensin,  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.  
 CC  
 SQ Sequence 1255 AA;  
 Query Match 99.0%; Score 6740; DB 23; Length 1255;  
 Best Local Similarity 99.0%; Pred. No. 0;  
 Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
 Qy 1 MELAALCRWGLLALLPGAASTOVCCTGDMKLRPASPEHLDMLRHLVQCGVYQGL 60  
 Db 1 MELAALCRWGLLALLPGAASTOVCCTGDMKLRPASPEHLDMLRHLVQCGVYQGL 60  
 Qy 61 ELTYLPTNASLFLQDIEVQGYVLAHNOVROVPLQRLRYRGTOLEEDNYALAVLNG 120  
 Db 61 ELTYLPTNASLFLQDIEVQGYVLAHNOVROVPLQRLRYRGTOLEEDNYALAVLNG 120  
 Qy 121 DPLNNTPTVTGASPGGLRELDRLSTELTKGVLQORNPOLCYOTIIMKOYIRANSKF 180  
 Db 121 DPLNNTPTVTGASPGGLRELDRLSTELTKGVLQORNPOLCYOTIIMKOYIRANSKF 180  
 Qy 181 IGITELNRSRACHPCSPCKGSRCKGSESEDCQSLTRIVCAGGCRKGPPLTDCHEQC 240  
 Db 181 LTLIDNRSRACHPCSPCKGSRCKGSESEDCQSLTRIVCAGGCRKGPPLTDCHEQC 240  
 Qy 241 AAGCTGPRGSDCLALFHNSGICELHCPALVTYNTDFESMPNDEGRTTFGASCVTACP 300  
 Db 241 AAGCTGPRGSDCLALFHNSGICELHCPALVTYNTDFESMPNDEGRTTFGASCVTACP 300  
 Qy 301 YNYLSTVGSCTVLCPLNNOVTAEDGTQRCBKSKPCARVYCYGIGMEHLEVRVATSAN 360  
 Db 301 YNYLSTVGSCTVLCPLNNOVTAEDGTQRCBKSKPCARVYCYGIGMEHLEVRVATSAN 360  
 Qy 361 IOEFAGCKIRGSLAFLEPESFDGPASNTAPLOPQLOVFTLEBITGYLYISAMPDSL 420  
 Db 361 IOEFAGCKIRGSLAFLEPESFDGPASNTAPLOPQLOVFTLEBITGYLYISAMPDSL 420  
 Qy 421 DLSVFQNLQVIRGRILHNHAYSLTLQIGISWLGRLSRLRELSGLALIHNTLHCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRILHNHAYSLTLQIGISWLGRLSRLRELSGLALIHNTLHCFVHTV 480  
 Qy 481 PMDQLFRNPHQALLHTANRPEDECVGEGIALCHQLCARHCHWPGFTQCVCNSQFLRGEC 540  
 Db 481 PMDQLFRNPHQALLHTANRPEDECVGEGIALCHQLCARHCHWPGFTQCVCNSQFLRGEC 540  
 Qy 541 VEEGRVLOGLPREVYNARHCLPCHPEOCPONGSVTCFPEADQCVACHYADPPCVAR 600  
 Db 541 VEEGRVLOGLPREVYNARHCLPCHPEOCPONGSVTCFPEADQCVACHYADPPCVAR 600  
 Qy 601 PSQVAPDLSYMPIMKFPDEBACOPCPINCHSCVDDDDKCPAERASPLTSIVSAVVG 660  
 Db 601 PSQVAPDLSYMPIMKFPDEBACOPCPINCHSCVDDDDKCPAERASPLTSIVSAVVG 660  
 Qy 661 ILVVVVGAVVFGILLIKRQOKIRRYTWRRLLQETELVPLTPSGAMPQOAMRIKETEL 720  
 Db 661 ILVVVVGAVVFGILLIKRQOKIRRYTWRRLLQETELVPLTPSGAMPQOAMRIKETEL 720  
 Qy 721 RKVAVLSGAGFTYKGIWIPDGENVKIPVAIKYLRNTSPKAKEILDEAYVMAVGSP 780  
 Db 721 RKVAVLSGAGFTYKGIWIPDGENVKIPVAIKYLRNTSPKAKEILDEAYVMAVGSP 780  
 Qy 781 YVSRLLGICLTSTVQVLTOLMPYGLLDHVENRGLSGODLLNMCMQIAKMSYLEDVR 840

```

Db      781  YVSRLLGICLTISTVLTOLMPYGCGLDHHVENRGRISQDLINMCQIAKMSYLEDR 840
Qy      841  LVHRDLAARNVLYKSPNHVKITDPGLARLIDIDETEHADGGKVPKMALESIIARRPT 900
Db      841  LVHRDLAARNVLYKSPNHVKITDPGLARLIDIDETEHADGGKVPKMALESIIARRPT 900
Qy      901  HOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVCKM 960
Db      901  HOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVCKM 960
Qy      961  IDSECRPRPRRELVSSESRMARDPORFVVI QNEDLGASPLDSTFYRLLEDMDGLVDA 1020
Db      961  IDSECRPRPRRELVSSESRMARDPORFVVI QNEDLGASPLDSTFYRLLEDMDGLVDA 1020
Qy      1021  EBYLPQOQGFCCPDAPAGAGVWVHRHRSSSTRSGGDLTLGLBSESEAPRSPPLASPG 1080
Db      1021  EBYLPQOQGFCCPDAPAGAGVWVHRHRSSSTRSGGDLTLGLBSESEAPRSPPLASPG 1080
Qy      1081  AGSDVFDGDLGMAKGLQSLPETHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db      1081  AGSDVFDGDLGMAKGLQSLPETHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy      1141  NOPDVRPOPSPRREBPPLPAARAGATLEBAKTLSPKKNVWVDFAFGAVENPEYLTPO 1200
Db      1141  NOPDVRPOPSPRREBPPLPAARAGATLEBAKTLSPKKNVWVDFAFGAVENPEYLTPO 1200
Qy      1201  GGAAPQHPPPAFSPAFDNLVYMDODPPERGA PPTSTFKGTPTAENPEYGLDVPV 1255
Db      1201  GGAAPQHPPPAFSPAFDNLVYMDODPPERGA PPTSTFKGTPTAENPEYGLDVPV 1255

RESULT 5
AA001111
ID  AA001111 standard; Protein; 1255 AA.
XX
AC  AA001111;
DT  01-JAN-1997 (first entry)
DE  HER-2/neu protein.
XX
KW  HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
    breast cancer; ovary cancer; colon cancer; lung cancer;
    prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS  Homo sapiens.
XX
FH  Key
FT  Domain
    Location/Qualifiers
    /label= "intracellular domain
    /note= "claimed domain, useful for immunisation"
XX
PN  WO9630514-A1.
PD  03-OCT-1996.
XX
PF  28-MAR-1996; 96WO-US01689.
PR  31-MAR-1995; 95US-0414417.
PA  (UNIW ) UNIV WASHINGTON.
PI  Cheever MA, Dieis MJ;
XX
DR  MPI; 1996-455361/45.
DR  N-PSDB; AAT40739.
XX
XX  DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
    treatment of malignancies with which the HER-2/neu oncogene is
    associated
XX
PS  Claim 2; Page 56-61; 71pp; English.

```

```

XX      Human HER-2/neu protein (AA001111), also called p185 or c-erbB2, is
CC      the product of the HER-2/neu oncogene (see also AAT40739). The
CC      protein is over-expressed in various cancers, including breast,
CC      ovarian, colon, lung and prostate. The intracellular domain of the
CC      protein can be used to immunise an animal against a malignancy with
CC      which the oncogene is associated. The polypeptide can be produced
CC      in transformed host cells for use in immunisation. Alternatively,
CC      animal cells are transfected in vivo or ex vivo with a viral vector
CC      that directs expression of the polypeptide.
XX
SQ      Sequence 1255 AA;

```

```

Query Match      98.9%; Score 6734; DB 17; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy      1  MELAALCRWGLLALLPGGAATGYCTGDMKRLRPASBETHLDMRLHYGGCQVVGNI 60
Db      1  MELAALCRWGLLALLPGGAATGYCTGDMKRLRPASBETHLDMRLHYGGCQVVGNI 60
Qy      61  ELTYLPTNASLSFLDDIQEVQCVLI IAHNQVQVPLQRLRIYRGTLQFEDNYALAVLDNG 120
Db      61  ELTYLPTNASLSFLDDIQEVQCVLI IAHNQVQVPLQRLRIYRGTLQFEDNYALAVLDNG 120
Qy      121  DPLNNTPTVAGSPGGLRELQRLSTEILKGGVLIQRNPQLCYODTITLWKDQYIKANSKF 180
Db      121  DPLNNTPTVAGSPGGLRELQRLSTEILKGGVLIQRNPQLCYODTITLWKDQYIKANSKF 180
Qy      181  IGITELNRSRACHPCSPMKCKGRGSESESDCOSLTRVYACAGCARCKPPLPTDCHEOC 240
Db      181  LTLLIDNRSRACHPCSPMKCKGRGSESESDCOSLTRVYACAGCARCKPPLPTDCHEOC 240
Qy      241  AAGCTGPRHSDCLACHFNHSGICELHCAALTYNTDTRESNPNBGRTRTFGASCYTACP 300
Db      241  AAGCTGPRHSDCLACHFNHSGICELHCAALTYNTDTRESNPNBGRTRTFGASCYTACP 300
Qy      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLQMEHLREAVATSAN 360
Db      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLQMEHLREAVATSAN 360
Qy      361  IOEPAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOQVFEETLEITGYLYISAMPDSL 420
Db      361  IOEPAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOQVFEETLEITGYLYISAMPDSL 420
Qy      421  DLSTVFQNLQVIRGRILHNGAVSLTLOGLISWLGRLSRLBELSGALIHNNHLCFVHTV 480
Db      421  DLSTVFQNLQVIRGRILHNGAVSLTLOGLISWLGRLSRLBELSGALIHNNHLCFVHTV 480
Qy      481  PMDQLFRNPQALHTANRPEDECVGEGLACHQLCARGHCMGPGPTQCVNCSOFLRGQEC 540
Db      481  PMDQLFRNPQALHTANRPEDECVGEGLACHQLCARGHCMGPGPTQCVNCSOFLRGQEC 540
Qy      541  VEECRVLQGLPREYVNAHRLCLPCHPECOFONGSVTCFGEADQCVAAHAYKDPFCVAC 600
Db      541  VEECRVLQGLPREYVNAHRLCLPCHPECOFONGSVTCFGEADQCVAAHAYKDPFCVAC 600
Qy      601  PSGVPRDLSYMPIMKFPDEEGACOPCPINCHSCVDLDKGPACORASPLTSIISAVVG 660
Db      601  PSGVPRDLSYMPIMKFPDEEGACOPCPINCHSCVDLDKGPACORASPLTSIISAVVG 660
Qy      661  ILVVVVLGVVFGIILKRRQOKIRKTYMRILQETELVEPLTPSGAMPNOAQRILKETEL 720
Db      661  ILVVVVLGVVFGIILKRRQOKIRKTYMRILQETELVEPLTPSGAMPNOAQRILKETEL 720
Qy      721  RKVKVLGSGAGCTGVYKGIWIPDGENVKI PVAIKVLRENTSPRANKELIDEAVVMGVGSP 780
Db      721  RKVKVLGSGAGCTGVYKGIWIPDGENVKI PVAIKVLRENTSPRANKELIDEAVVMGVGSP 780
Qy      781  YVSRLLGICLTISTVLTOLMPYGCGLDHHVENRGRISQDLINMCQIAKMSYLEDR 840
Db      781  YVSRLLGICLTISTVLTOLMPYGCGLDHHVENRGRISQDLINMCQIAKMSYLEDR 840

```

QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETETHADGKVP1KMMALLESILRRFT 900  
DB 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETETHADGKVP1KMMALLESILRRFT 900  
QY 901 HOSDVMASGYVTWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCM 960  
DB 901 HOSDVMASGYVTWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCM 960  
QY 961 IDSECRPRRELVSSESRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMDGLVDA 1020  
DB 961 IDSECRPRRELVSSESRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMDGLVDA 1020  
QY 1021 EETLVVQOGFCFCDDPARPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSP LAPSEG 1080  
DB 1021 EETLVVQOGFCFCDDPARPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSP LAPSEG 1080  
QY 1081 AASDVPDGDGLGMGAAGLQSLPTHDPSPLORYSEDPVPLPSETDGVAPLTCSPQPEYV 1140  
DB 1081 AASDVPDGDGLGMGAAGLQSLPTHDPSPLORYSEDPVPLPSETDGVAPLTCSPQPEYV 1140  
QY 1141 NOPDVARPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAPGGAVENTEYLTTPQ 1200  
DB 1141 NOPDVARPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAPGGAVENTEYLTTPQ 1200  
QY 1201 GGAAPQHPHPAPSPAFDNLVYWDOPPERGAPSPSTFKGTPTAENBEYLGLDVVPV 1255  
DB 1201 GGAAPQHPHPAPSPAFDNLVYWDOPPERGAPSPSTFKGTPTAENBEYLGLDVVPV 1255

RESULT 6  
AAM92406  
ID AAM92406 standard; Protein; 1255 AA.

AC AAM92406;  
DT 21-APR-1999 (first entry)  
XX Human HER-2/neu oncogene protein.  
DE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
KW malignancy; treatment; tumour.  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT Region 676..1255  
FT /note= "region which elicits immune response"  
PN US5869445-A.  
XX 09-FEB-1999.  
XX 01-APR-1996; 96US-0625101.  
XX 01-APR-1996; 96US-0625101.  
PR 17-MAR-1996; 96US-0625101.  
PR 12-AUG-1993; 93US-0033644.  
PR 31-MAR-1995; 95US-0414417.  
XX (UNIM ) UNIV WASHINGTON.  
PI Cheever MA, Disis ML;  
XX WPI; 1999-152835/13.  
DR N-PSDB; AAX01912.  
XX Use of HER-2/neu polypeptides - for eliciting an immune response to  
PT an HER-2/neu associated malignancy, particularly for treating or  
PT preventing tumours  
PS Claim 3; Column 31-38; 26pp; English.  
XX This sequence represents the human HER-2/neu oncogene protein. A fragment  
CC

CC of this protein is used in a method for eliciting or enhancing an immune  
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
CC B cells to produce an immune response to the HER-2/neu protein. The  
CC method can be used for immunisation against a malignancy in which the  
CC HER-2/neu oncogene is associated and in the treatment of an existing  
CC tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 20; Length 1255;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPFGAASQVCTGTDMLRLPASPEITHLMLHLYOGQVVGNTL 60  
DB 1 MELAALCRWGLLALLPFGAASQVCTGTDMLRLPASPEITHLMLHLYOGQVVGNTL 60  
QY 61 ELTYLPTNASLSLTDIOEYQGVYLAHNOVQVPLQRLRVGTQLFEDNVALLAVLDNG 120  
DB 61 ELTYLPTNASLSLTDIOEYQGVYLAHNOVQVPLQRLRVGTQLFEDNVALLAVLDNG 120  
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTIMKDOYIKANSKF 180  
DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTIMKDOYIKANSKF 180  
QY 181 IGTELNRSRACHPCSPMCKGSRCKWGESSEDCSLTRTVACAGGACARCKPPTDCCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMCKGSRCKWGESSEDCSLTRTVACAGGACARCKPPTDCCHEQC 240  
QY 241 AAGCTGPKKSDCLACIHFHNSGICELHCPALVTYNTDTFESMNPBGRYTFGASCVTACP 300  
DB 241 AAGCTGPKKSDCLACIHFHNSGICELHCPALVTYNTDTFESMNPBGRYTFGASCVTACP 300  
QY 301 YNYLSTDVSGCTLVCPHNOEVTAEEDTORCEKCSRPCARVCYGLGMEHLREYRAVTSAN 360  
DB 301 YNYLSTDVSGCTLVCPHNOEVTAEEDTORCEKCSRPCARVCYGLGMEHLREYRAVTSAN 360  
QY 361 IOEFACCKKIFGSLAPESPFDGDPASNTAPLOPELOVETLEETIGLYISAMPDLP 420  
DB 361 IOEFACCKKIFGSLAPESPFDGDPASNTAPLOPELOVETLEETIGLYISAMPDLP 420  
QY 421 DLSVFONLOVIRGRIIANGAYSLTLOGLISWLGRLSRLSGLALIHNTHLCPVHTV 480  
DB 421 DLSVFONLOVIRGRIIANGAYSLTLOGLISWLGRLSRLSGLALIHNTHLCPVHTV 480  
QY 481 PMDQLFRNPHOALHTANRPEDECVGEGLACHOLCARGCMGSGPQCVNCSQFLRGQEC 540  
DB 481 PMDQLFRNPHOALHTANRPEDECVGEGLACHOLCARGCMGSGPQCVNCSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
QY 601 PSGVKPDLSTYMPIWKFPEDEGACQPCPINCTHSCVDLDDKCPAEORASPLTISVAVVG 660  
DB 601 PSGVKPDLSTYMPIWKFPEDEGACQPCPINCTHSCVDLDDKCPAEORASPLTISVAVVG 660  
QY 661 ILLVVLGVVFGILLIRROOKIRKTYMRLLQETELVEPLTPBGAAMPNOAMILKETEL 720  
DB 661 ILLVVLGVVFGILLIRROOKIRKTYMRLLQETELVEPLTPBGAAMPNOAMILKETEL 720  
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKEILDEAYMAGVSP 780  
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKEILDEAYMAGVSP 780  
QY 781 YVSRILGICLTSTVOQLVTLQMPYGLLDHVRENRGLSGQDLLNMCQIAKGSYLEDYR 840  
DB 781 YVSRILGICLTSTVOQLVTLQMPYGLLDHVRENRGLSGQDLLNMCQIAKGSYLEDYR 840  
QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETETHADGKVP1KMMALLESILRRFT 900  
DB 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETETHADGKVP1KMMALLESILRRFT 900

```

QY 901 HOSDWSYGVTVLMTFGAKPYDGI PAREI PDLLEKGERLPQPICTIDVYIMVKCM 960
DB 901 HOSDWSYGVTVLMTFGAKPYDGI PAREI PDLLEKGERLPQPICTIDVYIMVKCM 960
QY 961 IDSECRPRFRELSEFSRMA RDPQRFV IONEDLGPA SPLDSTFYRS LLEDDMGDLVDA 1020
DB 961 IDSECRPRFRELSEFSRMA RDPQRFV IONEDLGPA SPLDSTFYRS LLEDDMGDLVDA 1020
QY 1021 EELVLVQOQGFCCDPAPGAGCMVHHRRSSSTRSGGDLTLGLPSESEAPRSPAPSEG 1080
DB 1021 EELVLVQOQGFCCDPAPGAGCMVHHRRSSSTRSGGDLTLGLPSESEAPRSPAPSEG 1080
QY 1081 AGSDVDFDGLGMAAKLQSLPTHDPSPLO RYSEDPVPLPSETDGVAVLTCSPQPEYV 1140
DB 1081 AGSDVDFDGLGMAAKLQSLPTHDPSPLO RYSEDPVPLPSETDGVAVLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPA GATLERAKTISPGKNGVVKDVFAGAVENPEYLTPO 1200
DB 1141 NOPDVRPQPPSPREGPLPAARPA GATLERAKTISPGKNGVVKDVFAGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPSPAFNDLYYWDODPBERGAPSTFKGPTAENPEYLGLDVPU 1255
DB 1201 GGAAPQHPPPAPSPAFNDLYYWDODPBERGAPSTFKGPTAENPEYLGLDVPU 1255

```

RESULT 7  
AAB21198 standard; protein; 1255 AA.

AAB21198;

12-JAN-2001 (first entry)

Human HER-2/neu protein.

Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
breast cancer; prostate cancer; ovarian cancer; lung cancer;

colon cancer.

Homo sapiens.

MO200044899-A1.

03-AUG-2000.

28-JAN-2000; 2000MO-US02164.

29-JAN-1999; 99US-0117976.

(COR1-) CORIXA CORP.

(SMIK) SMITHKLINE BEECHAM.

Cheever MA, Gheysen D;

WPI; 2000-505976/45.

N-PSDB; AAA89736.

HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
useful for vaccinating against breast, ovarian, colon, lung and

prostate cancers -

Claim 52; Fig 7; 128pp; English.

The present sequence is the human HER-2/neu protein. It is a member of  
the tyrosine kinase family of receptor-like glycoproteins and shows  
homology to the epidermal growth factor receptor (EGFR). It probably  
plays a part in cell growth and/or differentiation. The HER-2/neu  
gene is an oncogene. An HER-2/neu fusion protein comprising a  
HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
domain may be used to treat or prevent cancer by eliciting or  
enhancing an immune response to the HER-2/neu protein. It may be used  
to treat malignancies such as breast, ovarian, colon, lung and  
prostate cancers, and may be used as an antigen to vaccinate against

```

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 98.9%; Score 6734; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 MELAALCRMGILLALLPAGAASQVCTGDMKRLPASPTHTDMLRHLYOGGCVVGNL 60
DB 1 MELAALCRMGILLALLPAGAASQVCTGDMKRLPASPTHTDMLRHLYOGGCVVGNL 60
QY 61 ELLYLTNMSLSFLQDIQEOGVVLI AHNOVRQVPLRLRI VNGTQLFEDNVLAVALDNG 120
DB 61 ELLYLTNMSLSFLQDIQEOGVVLI AHNOVRQVPLRLRI VNGTQLFEDNVLAVALDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQSLTEILKGVLIQRNPOLCYQDITLMDQYIKANSKF 180
DB 121 DPLNNTPTVTGASPGGLRELQSLTEILKGVLIQRNPOLCYQDITLMDQYIKANSKF 180
QY 181 IGTTELNRSPACHPCSPMCKSGRCWGBSSSDCQSLRTTVAGAGCARCKGPLTDCCHEOC 240
DB 181 IGTTELNRSPACHPCSPMCKSGRCWGBSSSDCQSLRTTVAGAGCARCKGPLTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVAVTSAN 360
QY 361 IOEFACCKKIFGSLAPLPESEFDGPASNTAPLOPELOQVETLEETITGVYISAMPDLP 420
DB 361 IOEFACCKKIFGSLAPLPESEFDGPASNTAPLOPELOQVETLEETITGVYISAMPDLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLSLRSLRSGALIIHNTHLCFHYTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLSLRSLRSGALIIHNTHLCFHYTV 480
QY 481 PMDQLFRNPHQALLHTANRDEDECVEBGLACHOLCARGHCWGQPTQCVNCSQFLRGQC 540
DB 481 PMDQLFRNPHQALLHTANRDEDECVEBGLACHOLCARGHCWGQPTQCVNCSQFLRGQC 540
QY 541 VEECRVLOGLPREYVNAHRCLPCHPECPQNGSVTCGPBADCVACAHRKDPFCVARC 600
DB 541 VEECRVLOGLPREYVNAHRCLPCHPECPQNGSVTCGPBADCVACAHRKDPFCVARC 600
QY 601 PSGVKPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDDKGCFAEGRASPLTGISAVNG 660
DB 601 PSGVKPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDDKGCFAEGRASPLTGISAVNG 660
QY 661 ILLVVLGVVFGILIKRRQOKIRKYTRRLLQSTELVEPLTPSGAMPNOQMRLKETEL 720
DB 661 ILLVVLGVVFGILIKRRQOKIRKYTRRLLQSTELVEPLTPSGAMPNOQMRLKETEL 720
QY 721 RKRYKLGSGAFGVNGIWI PDGENVKIPIAIVLLENTSPKANKELIDDAVYMAAGGSP 780
DB 721 RKRYKLGSGAFGVNGIWI PDGENVKIPIAIVLLENTSPKANKELIDDAVYMAAGGSP 780
QY 781 YVSRLLGICLTSTVOLQMPYGCLLDHYRENGRGLGSDLLNMCMQIKAGNSYLEDVR 840
DB 781 YVSRLLGICLTSTVOLQMPYGCLLDHYRENGRGLGSDLLNMCMQIKAGNSYLEDVR 840
QY 841 LVHRDLAARVNLVKSPPNHVKTDFGLARLLDIDETEVHADGKVPKIMWLESILRRFT 900
DB 841 LVHRDLAARVNLVKSPPNHVKTDFGLARLLDIDETEVHADGKVPKIMWLESILRRFT 900
QY 901 HOSDWSYGVTVLMTFGAKPYDGI PAREI PDLLEKGERLPQPICTIDVYIMVKCM 960
DB 901 HOSDWSYGVTVLMTFGAKPYDGI PAREI PDLLEKGERLPQPICTIDVYIMVKCM 960
QY 961 IDSECRPRFRELSEFSRMA RDPQRFV IONEDLGPA SPLDSTFYRS LLEDDMGDLVDA 1020

```

Db 961 IDSCRRPREFELVSEFSRMARDPQRFVVIQNEDELGPAPPLDSTFYRSLLEDMDGLVDA 1020  
Qy 1021 EEVLPVPOGFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG 1080  
Db 1021 EEVLPVPOGFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG 1080  
Qy 1081 AGSDVFDGDLQMGAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPOPEYV 1140  
Db 1081 AGSDVFDGDLQMGAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPOPEYV 1140  
Qy 1141 NOPVRPQPSRPGPLPAPRPAATLERAKTLSPGKGVKDVAFGAVENPEYLTPQ 1200  
Db 1141 NOPVRPQPSRPGPLPAPRPAATLERAKTLSPGKGVKDVAFGAVENPEYLTPQ 1200  
Qy 1201 GGAAPQHPHPAPSPAFDNLTYWDDPPERGAPSTFKGPTTAENPEYLGLDVPY 1255  
Db 1201 GGAAPQHPHPAPSPAFDNLTYWDDPPERGAPSTFKGPTTAENPEYLGLDVPY 1255  
RESULT 8  
AAV84780 ID AAV84780 standard; Protein; 1255 AA.  
AC AAV84780;  
DT 08-AUG-2000 (first entry)  
DE Amino acid sequence of the SPLICE erdb-2 receptor protein.  
KM SPLICE erdb-2 receptor protein; cell transformation disorder; cancer;  
KW tumor cell proliferation; tissue degeneration; arthropathy;  
KM bone resorption; inflammatory disease; degenerative disorder;  
KM wound healing.  
OS Homo sapiens.  
XX W0200020579-A1.  
PN 13-APR-2000.  
PD 13-APR-2000.  
XX 01-OCT-1999; 99WO-CA00912.  
PF 01-OCT-1999; 99WO-CA00912.  
PR 02-OCT-1998; 98US-0165192.  
XX (UYMC-) UNIV MCMMASTER.  
PA Muller WJ, Siegel PM;  
PI WPI; 2000-303768/26.  
DR N-PSDB; AAA14812.  
XX Nucleic acid encoding an erdb 2 receptor protein designated SPLICE  
PT erdb-2, inhibitors of the protein are useful for treatment of cancer -  
PS Claim 3; Fig 2; 60pp; English.  
XX The present sequence represents a SPLICE erdb-2 receptor protein. The  
CC protein has an in-frame deletion of 16 amino acids, 2 of which are  
CC conserved cysteine residues, compared to the unspliced protein. The  
CC erdb-2 polynucleotide is used to construct probes for detecting  
CC disorders of cell transformation such as cancer. Antibodies to the  
CC protein may be used to detect SPLICE erdb-2 in a sample. Agents  
CC (e.g. antisense oligonucleotides) which inhibit the expression of  
CC SPLICE erdb-2 are useful for reducing tumor cell proliferation and  
CC treating cancer. Substances which stimulate SPLICE erdb-2 are useful  
CC in which degeneration of tissue occurs, such as arthropathy, bone  
CC resorption, inflammatory diseases, degenerative disorders of the  
CC central nervous system and wound healing.  
XX Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 21; Length 1255;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
Qy 1 MELAALCWGILLALLPPGAASVCTCTGDKLRLPAPRETHLMDRLYLQGCQVQGNL 60  
Db 1 MELAALCWGILLALLPPGAASVCTCTGDKLRLPAPRETHLMDRLYLQGCQVQGNL 60  
Qy 61 ELTYLPTVASLSFLQDIOEVQGYVLIANOVQVQLQRLIRVGTQLFEDNALVLNDG 120  
Db 61 ELTYLPTVASLSFLQDIOEVQGYVLIANOVQVQLQRLIRVGTQLFEDNALVLNDG 120  
Qy 121 DPLNNTPTVGTASPGGLRELQRLSTEILKGVLIQRNPOLCYODTILMKOYIKANSKF 180  
Db 121 DPLNNTPTVGTASPGGLRELQRLSTEILKGVLIQRNPOLCYODTILMKOYIKANSKF 180  
Qy 121 DPLNNTPTVGTASPGGLRELQRLSTEILKGVLIQRNPOLCYODTILMKOYIKANSKF 180  
Db 121 DPLNNTPTVGTASPGGLRELQRLSTEILKGVLIQRNPOLCYODTILMKOYIKANSKF 180  
Qy 181 IGITELIRSRACHCSPPCKSKSCRCGSESEDCQSITRVCAGGCARCGPLTDCCHQC 240  
Db 181 LTLIDYRSRACHCSPPCKSKSCRCGSESEDCQSITRVCAGGCARCGPLTDCCHQC 240  
Qy 241 AAGCTGPHGSDCLCLPHNHSIGICELHCPALVTYNTDFEEMPNPEGRYTGCACVTCAP 300  
Db 241 AAGCTGPHGSDCLCLPHNHSIGICELHCPALVTYNTDFEEMPNPEGRYTGCACVTCAP 300  
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRATVSAN 360  
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRATVSAN 360  
Qy 361 IQEFAGCKKRTGSALFPESEFDGDPASNTAPLQEPOLQVFEETLEITGYLIASMPDLP 420  
Db 361 IQEFAGCKKRTGSALFPESEFDGDPASNTAPLQEPOLQVFEETLEITGYLIASMPDLP 420  
Qy 421 DLSVFQNLQVIRGRIIHNHAYSLTLOGISLWLSRLRELSGLALIHNTHTLCFYHTV 480  
Db 421 DLSVFQNLQVIRGRIIHNHAYSLTLOGISLWLSRLRELSGLALIHNTHTLCFYHTV 480  
Qy 481 PMDQLFRNPQALHTANRPEDECBGGLACHOLCARGHCGPPTQCVNCSQFLRGQEC 540  
Db 481 PMDQLFRNPQALHTANRPEDECBGGLACHOLCARGHCGPPTQCVNCSQFLRGQEC 540  
Qy 541 VEECRVUQGLPREVYNAHCLPCHPECOPONGSITCFEPADQCYACHYDPPFCVARC 600  
Db 541 VEECRVUQGLPREVYNAHCLPCHPECOPONGSITCFEPADQCYACHYDPPFCVARC 600  
Qy 601 PSQVPPDLSYMPKIPFDEBGACOPCPINCHSCVDLDDKCPAQRASPLTSTVSAYG 660  
Db 601 PSQVPPDLSYMPKIPFDEBGACOPCPINCHSCVDLDDKCPAQRASPLTSTVSAYG 660  
Qy 661 ILVVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPQAOAMRLKETEL 720  
Db 661 ILVVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPQAOAMRLKETEL 720  
Qy 721 RKXVVLGSAFGTYKGIWIPDGENVKI PVAIKYLRRENTSPKANKIIDEAYVAVGSP 780  
Db 721 RKXVVLGSAFGTYKGIWIPDGENVKI PVAIKYLRRENTSPKANKIIDEAYVAVGSP 780  
Qy 781 YVSRLLGICLTSTVOLTOLMPYGCILDHRENRGRGSDLLMWQOIAAGNSYLEDVR 840  
Db 781 YVSRLLGICLTSTVOLTOLMPYGCILDHRENRGRGSDLLMWQOIAAGNSYLEDVR 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLIDIBETEHADGAKVPIKMALESILRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLIDIBETEHADGAKVPIKMALESILRRFT 900  
Qy 901 HOSDWSYGYTVELMTFGAKPYDGIIPAREIPDLERGERLPORPCTIDIVYIMVXCM 960  
Db 901 HOSDWSYGYTVELMTFGAKPYDGIIPAREIPDLERGERLPORPCTIDIVYIMVXCM 960  
Qy 961 IDSCRRPREFELVSEFSRMARDPQRFVVIQNEDELGPAPPLDSTFYRSLLEDMDGLVDA 1020  
Db 961 IDSCRRPREFELVSEFSRMARDPQRFVVIQNEDELGPAPPLDSTFYRSLLEDMDGLVDA 1020  
Qy 1021 EEVLPVPOGFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG 1080



Db 1021 EEYLVPOGGFCPCDPAPGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSG 1080  
Qy 1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEV 1140  
Db 1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEV 1140  
Qy 1141 NOPDVRPOPSPREGPLPAARPAAGATLEERAKTSLPGKNGVVDVFAFGAVENPEYLTPO 1200  
Db 1141 NOPDVRPOPSPREGPLPAARPAAGATLEERAKTSLPGKNGVVDVFAFGAVENPEYLTPO 1200  
Qy 1201 GGAAQPPHPPAPFAFDMLYYWDDPPERGAPESTFKGTPTAENPEYLGIDVPV 1255  
Db 1201 GGAAQPPHPPAPFAFDMLYYWDDPPERGAPESTFKGTPTAENPEYLGIDVPV 1255

## RESULT 9

AAB85458  
ID AAB85458 standard; Protein; 1255 AA.

AC AAB85458;

DT 25-SEP-2001 (first entry)

DE Human HER-2/neu protein.

KM Antigen-presenting cell; immunogenic; immune response; HER-2/neu; oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.

OS Homo sapiens.

PN WO200153463-A2.

PD 26-JUL-2001.

PF 19-JAN-2001; 2001WO-US01850.

PR 21-JAN-2000; 2000US-0177545.

PA (CORI-) CORIXA CORP.

PI Cheever MA, Hand-Zimmermann S;

DR WPI; 2001-476112/51.

DR N-PSDB; AAH23392.

PT New antigen-presenting cells, useful as vaccines for eliciting or enhancing an immune response to HER-2/neu protein, particularly useful for treating or preventing cancer, e.g. breast cancer -

PS Claim 2; Page 41-46; 49pp; English.

CC The invention provides an isolated antigen-presenting cell, which expresses at least an immunogenic portion of a polypeptide that produces an immune response to HER-2/neu protein. The antigen-presenting cells are useful as vaccines for eliciting or enhancing an immune response to HER-2/neu protein, particularly in treating or preventing malignancies in which the HER-2/neu oncogene is associated. Specifically, these are useful for treating or preventing cancer, e.g. breast cancer, ovarian, colon, lung or prostate cancers. The present sequence represents the human HER-2/neu protein (also known as p185 or c-erbB2).

CC Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 22; Length 1255;

Best Local Similarity 98.8%; Fred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MELALACRWGLLALPLPGASTOVCTGDMKLRPLPASPETHLDMRLHYOGGQVYQGM 60  
Db 1 MELALACRWGLLALPLPGASTOVCTGDMKLRPLPASPETHLDMRLHYOGGQVYQGM 60  
Qy 61 ELTYLPTNASLSFLDIOEVGGVLIANHQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLDIOEVGGVLIANHQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
Qy 121 DPLANTTPVTGASPGGLRELQRLSLTEILKGVLIORNPQLCYODTILKDOYIKANSKF 180  
Db 121 DPLANTTPVTGASPGGLRELQRLSLTEILKGVLIORNPQLCYODTILKDOYIKANSKF 180  
Qy 181 IGTETLNSRACHPCSPMKSGRCWGESESDQSILRTYACAGCANKCPPLPDCHEOC 240  
Db 181 IGTETLNSRACHPCSPMKSGRCWGESESDQSILRTYACAGCANKCPPLPDCHEOC 240  
Qy 241 AAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTESPNPEGRYTFGASCTYAC 300  
Db 241 AAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTESPNPEGRYTFGASCTYAC 300  
Qy 301 YNYLSTDVGSCTLVCPRLHNOEYTABDGTORCEKSKPCARVCYGLAMEHLREYRAVTSAN 360  
Db 301 YNYLSTDVGSCTLVCPRLHNOEYTABDGTORCEKSKPCARVCYGLAMEHLREYRAVTSAN 360  
Qy 361 IOEPAGCKKI FGSILAFPLESPDGPASNTAPLOPELOVFEETLEETGYLYISAMPDSL 420  
Db 361 IOEPAGCKKI FGSILAFPLESPDGPASNTAPLOPELOVFEETLEETGYLYISAMPDSL 420  
Qy 421 DLSVFQNLQVIRGRIILHNGAVSLTLQGLISWLGRLSRLSGGLALIHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRIILHNGAVSLTLQGLISWLGRLSRLSGGLALIHNNHLCFVHTV 480  
Qy 481 PMDQLFRNHQALLHTANRPEDECVGEGLAQHLCARHGWGPGPTOCVNCQFLRGQEC 540  
Db 481 PMDQLFRNHQALLHTANRPEDECVGEGLAQHLCARHGWGPGPTOCVNCQFLRGQEC 540  
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECOPONSVCYCFEADQCVACAHYKOPFCVAC 600  
Db 541 VEECRVLOGLPREYVNAHCLPCHPECOPONSVCYCFEADQCVACAHYKOPFCVAC 600  
Qy 601 PSGVAPDLSYMPIMKFPDEEGACOPCINCTHSCVDLDKGCPAEORASPLTISAVVG 660  
Db 601 PSGVAPDLSYMPIMKFPDEEGACOPCINCTHSCVDLDKGCPAEORASPLTISAVVG 660  
Qy 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOAKMILKETEL 720  
Db 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOAKMILKETEL 720  
Qy 721 RKVKVLGSGAFCTVYKGIPIPDGENVKIPVALIKVLRENTSPRANKETLDEAYMAVGSP 780  
Db 721 RKVKVLGSGAFCTVYKGIPIPDGENVKIPVALIKVLRENTSPRANKETLDEAYMAVGSP 780  
Qy 781 YVSRLLGICLTSTVQLVQMLPYGCLLDHVRNCRGLSGODLLNMCMQIAKMSYLEDR 840  
Db 781 YVSRLLGICLTSTVQLVQMLPYGCLLDHVRNCRGLSGODLLNMCMQIAKMSYLEDR 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETERYHADGKVPKIMALESILRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETERYHADGKVPKIMALESILRRFT 900  
Qy 901 HOSDWSGVYTWELMTREKAKKYDGI PAEIEDLLEKGRLOPPCTIDVYIMVKCM 960  
Db 901 HOSDWSGVYTWELMTREKAKKYDGI PAEIEDLLEKGRLOPPCTIDVYIMVKCM 960  
Qy 961 IDSECRPFRELVSFSRMAPRQFVVYQNEIDLPAASLDSTFRSLLEDMDGDLVYA 1020  
Db 961 IDSECRPFRELVSFSRMAPRQFVVYQNEIDLPAASLDSTFRSLLEDMDGDLVYA 1020  
Qy 1021 EEYLVPOGGFCPCDPAPGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSG 1080  
Db 1021 EEYLVPOGGFCPCDPAPGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSG 1080  
Qy 1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEV 1140  
Db 1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEV 1140  
Qy 1141 NOPDVRPOPSPREGPLPAARPAAGATLEERAKTSLPGKNGVVDVFAFGAVENPEYLTPO 1200



Db 1141 NOPDVRPDPSPBEGPLPAAPAGATLERPKTLSPKNGVVKDVFAGAVENPEYLTPQ 1200  
QY 1201 GGAAPQHPHPAFSPAFDNLYYWDODPPERGAPSPSTFKGTPTAENPEYLGIDVPV 1255  
DB 1201 GGAAPQHPHPAFSPAFDNLYYWDODPPERGAPSPSTFKGTPTAENPEYLGIDVPV 1255  
RESULT 10  
AAG88267  
ID AAG88267 standard; Protein; 1255 AA.  
XX AAG88267;  
AC AAG88267;  
XX 11-SEP-2001 (first entry)  
DT 11-SEP-2001 (first entry)  
XX HER2/neu amino acid sequence.  
DE HER2/neu amino acid sequence.  
XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
KW Immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;  
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO200141787-A1.  
PN 14-JUN-2001.  
PD 14-JUN-2001.  
XX 11-DEC-2000; 2000WO-US33591.  
PF 11-DEC-2000; 2000WO-US33591.  
XX 10-DEC-1999; 99US-0458299.  
PR 10-DEC-1999; 99US-0458299.  
XX (EPIIM-) EPIIMCNE INC.  
PA Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E,  
PI Keogh E;  
PI WPI: 2001-374995/39.  
DR WPI: 2001-374995/39.  
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
PT cellular immune responses for the prevention and treatment of cancer -  
XX  
XX Disclosure: Page 15; 1999p; English.  
XX  
XX The present invention describes isolated prepared HER2/neu epitopes (I).  
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
CC culture in vitro and binds to a complex of an epitope (II), bound to a  
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)  
CC and a second epitope and the peptide is less than 50 contiguous amino  
CC acids that have 100% identity with a native peptide sequence of HER2/neu;  
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising  
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic  
CC and immunostimulant activities, and can be used in vaccines. (I), (II)  
CC and (III) are useful for inducing cellular immune responses for the  
CC prevention and treatment of cancer. (I) and (II) are useful for  
CC monitoring or evaluating an immune response to a tumour-associated  
CC antigen when incubated with a T lymphocyte sample from a patient and  
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
CC based vaccines mean that immunosuppressive epitopes that may be present  
CC in whole antigens may be avoided. Selected epitopes may be combined to  
CC enhance immunogenicity. The possible pathological side effects caused by  
CC infectious agents or whole protein antigen is eliminated. The vaccine  
CC provides the ability to direct and focus an immune response to multiple  
CC selected antigens from the same pathogen. Epitope-based anti-tumour  
CC vaccines provides the opportunity to combine epitopes derived from  
CC multiple tumour-associated molecules addressing the problem of tumour-  
CC tumour variability and reducing the likelihood of tumour escape due to  
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
CC the exemplification of the present invention.  
XX  
XX Sequence 1255 AA;  
SQ

Query Match 98.9%; Score 6734; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLTLLALPPAASTOYCTGDMTLRLPASFTLMDMLRHLYOCGVVQGNL 60  
DB 1 MELAALCRWGLTLLALPPAASTOYCTGDMTLRLPASFTLMDMLRHLYOCGVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIEVQGVYLIANHQVQVPLQRIYRGTOLEFENYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIEVQGVYLIANHQVQVPLQRIYRGTOLEFENYALAVLDNG 120  
QY 121 DPLNNTTPTVGASPGGLRELOLRSLTEILKGVVLIQRNOLCYODTILMKDVIKANSKF 180  
DB 121 DPLNNTTPTVGASPGGLRELOLRSLTEILKGVVLIQRNOLCYODTILMKDVIKANSKF 180  
QY 181 IGITELNLSRACHPSSPMCKSGRCGSESEDDQSLTRYTCAGGCARCKRPLPDDCCEOC 240  
DB 181 LTLIDTNSRACHPSSPMCKSGRCGSESEDDQSLTRYTCAGGCARCKRPLPDDCCEOC 240  
QY 241 AAGCTGPKHSDCLAFHNSGICELHCPALVTYNTDTPESMPNPEGRTTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLAFHNSGICELHCPALVTYNTDTPESMPNPEGRTTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLEHNOEVTABDGTORCEKSKPCARVCYGLMEHLREVAVTSAN 360  
DB 301 YNYLSTDVGSCTLVCPLEHNOEVTABDGTORCEKSKPCARVCYGLMEHLREVAVTSAN 360  
QY 361 IOEFAGCKKIRGSLAFIPESPFGDPASNTAPIQEPOLQVFELEITGLYISAMPDSLP 420  
DB 361 IOEFAGCKKIRGSLAFIPESPFGDPASNTAPIQEPOLQVFELEITGLYISAMPDSLP 420  
QY 421 DLSVFQNTQVLRGRILHNGAVSLTLQGISMLGLRSLRELGSLALIHNNHLCVPHIV 480  
DB 421 DLSVFQNTQVLRGRILHNGAVSLTLQGISMLGLRSLRELGSLALIHNNHLCVPHIV 480  
QY 481 PWDOLFERNPHOALLTANRPEDECVGEGILACGHCARHCMPGPTOCVNCSCSFLRGQEC 540  
DB 481 PWDOLFERNPHOALLTANRPEDECVGEGILACGHCARHCMPGPTOCVNCSCSFLRGQEC 540  
QY 541 VEECVVLOGLPREYVNAHCLPCHPECCOPONGSVTCFGEADQCAVCAHYKDPFCVAPC 600  
DB 541 VEECVVLOGLPREYVNAHCLPCHPECCOPONGSVTCFGEADQCAVCAHYKDPFCVAPC 600  
QY 601 PSGVPELDSYPIWPFPEBEGACOPCINCHSCVDLDDKGPAPORASPLTSYSAVVG 660  
DB 601 PSGVPELDSYPIWPFPEBEGACOPCINCHSCVDLDDKGPAPORASPLTSYSAVVG 660  
QY 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOQMRILKETEL 720  
DB 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOQMRILKETEL 720  
QY 721 RKVKYLSGAGTGYKGIWIDGENVKI PVAIKVIRENTSPRANKEILDEAVYMGVGP 780  
DB 721 RKVKYLSGAGTGYKGIWIDGENVKI PVAIKVIRENTSPRANKEILDEAVYMGVGP 780  
QY 781 YVSRLLGICLSTYQVLYQMLPYGCLLDHVENRRLSDOLNMCQIAKMSYLEVDR 840  
DB 781 YVSRLLGICLSTYQVLYQMLPYGCLLDHVENRRLSDOLNMCQIAKMSYLEVDR 840  
QY 841 LVHRDLAARNLVKSPNHTKITDGLARLIDETEHADGKVPYIKMALESILRRFT 900  
DB 841 LVHRDLAARNLVKSPNHTKITDGLARLIDETEHADGKVPYIKMALESILRRFT 900  
QY 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKBERLPQPICTIDVYMIWVKCM 960  
DB 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKBERLPQPICTIDVYMIWVKCM 960  
QY 961 IDSCRPFRRLVSEFSMARDDPQRFVYIQMEDJGPAJPLUSTPFRSLLEDDDKDLYDA 1020  
DB 961 IDSCRPFRRLVSEFSMARDDPQRFVYIQMEDJGPAJPLUSTPFRSLLEDDDKDLYDA 1020  
QY 1021 EBYLVPOGFFCPDPAAGGMVHRHRSSTRSGGDLTLGLSPSEEARSPPLAPSEG 1080  
DB 1021 EBYLVPOGFFCPDPAAGGMVHRHRSSTRSGGDLTLGLSPSEEARSPPLAPSEG 1080

Dh 1021 EEYLVPQOGFCPCDPAPGAGVHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080  
Qy 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
Dh 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
Qy 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200  
Dh 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200  
Qy 1201 GGAAPOPHPPAPSPAFNDLYYWDODPPERGAPSTFKGPTTANENPEYLGIDVPV 1255  
Dh 1201 GGAAPOPHPPAPSPAFNDLYYWDODPPERGAPSTFKGPTTANENPEYLGIDVPV 1255

## RESULT 11

AAE24067

ID AAE24067 standard; Protein; 1255 AA.

AAE24067;

DT 23-SEP-2002 (first entry)

DE Human Her-2 protein.

KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;  
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;  
KW tumour; gene therapy; phosphorothioate backbone.

OS Homo sapiens.

PN WO200222636-A1.

PD 21-MAR-2002.

PF 12-SEP-2001; 2001WO-US28572.

PR 15-SEP-2000; 2000US-0663834.

PA (ISIS-) ISIS PHARM INC.

PI Bennett CF, Coweert LM;

DR WPI; 2002-471192/50.

DR N-PSDB; AAD38904.

PT Novel antisense oligonucleotide which modulates the expression of Human  
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors  
PT Inflammation or to prevent infection in humans -

PS Example 13; Page 95-107; 116pp; English.

CC The invention relates to antisense compounds targeted to a nucleic  
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)  
CC that specifically hybridises with and inhibits the expression of Her2.  
CC Antisense compounds of the invention are used for treating diseases or  
CC conditions associated with Her2 such as hyperproliferative disorders  
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,  
CC neural or cardiac cancer. They are also useful prophylactically e.g.  
CC to prevent or delay infection, inflammation and tumour formation. The  
CC invention is also used in gene therapy. The present sequence is human  
CC Her-2 protein.

SQ Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 23; Length 1255;

Beet Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MEIAALCRWGLTALTPGAASTOVCTGTDMLRLPASPTHMLRLHYOGGVQGNL 60  
Dh 1 MEIAALCRWGLTALTPGAASTOVCTGTDMLRLPASPTHMLRLHYOGGVQGNL 60

Qy 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200

Qy 61 ELTYLPTNALSFLQDIOEVQGVLLIAHNOVRQVPLQRLRIVRGTOLFEEDNVALAVLNG 120  
Dh 61 ELTYLPTNALSFLQDIOEVQGVLLIAHNOVRQVPLQRLRIVRGTOLFEEDNVALAVLNG 120  
Qy 121 DPLNNTPTVTGASPGGLRELQRLSTLEILLKGVLIQRNPOLCYODTILMKDOYIKANSKF 180  
Dh 121 DPLNNTPTVTGASPGGLRELQRLSTLEILLKGVLIQRNPOLCYODTILMKDOYIKANSKF 180  
Qy 181 IGITELNRACHPCSPMKSGRCSGSESSDDQSLTRTYCAGGACCKPPLPDCCHEOC 240  
Dh 181 IGITELNRACHPCSPMKSGRCSGSESSDDQSLTRTYCAGGACCKPPLPDCCHEOC 240  
Qy 241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTESMNPREGRYFFGASCVTACP 300  
Dh 241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTESMNPREGRYFFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEEDGTORCEKSPKPCARVCYGLMEHLREVRVTSAN 360  
Dh 301 YNYLSTDVGSCTLVCPHLNQEVTAEEDGTORCEKSPKPCARVCYGLMEHLREVRVTSAN 360  
Qy 361 IOEFAGCKKIFGSLAFSPSGDPAASNTAPLQEQLOVEFTELEITGYLYISAMPDLSL 420  
Dh 361 IOEFAGCKKIFGSLAFSPSGDPAASNTAPLQEQLOVEFTELEITGYLYISAMPDLSL 420  
Qy 421 DLSVFQNLQVIRGRIILHNGAYSLTQGLGISWLGRLSRLSGSLALIHNTHLCEVHTV 480  
Dh 421 DLSVFQNLQVIRGRIILHNGAYSLTQGLGISWLGRLSRLSGSLALIHNTHLCEVHTV 480  
Qy 481 PMDQLFRNPQALHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
Dh 481 PMDQLFRNPQALHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECOQONSVTTCFGEADOCVACAHKDPFCVARC 600  
Dh 541 VEECRVLOGLPREYVNAHCLPCHPECOQONSVTTCFGEADOCVACAHKDPFCVARC 600  
Qy 601 PSGVKPDLSTYMPIMKPEDEGACQPCPINCTHSCVDLDDKGCFAEGRASPSTISAVVG 660  
Dh 601 PSGVKPDLSTYMPIMKPEDEGACQPCPINCTHSCVDLDDKGCFAEGRASPSTISAVVG 660  
Qy 661 ILLVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOQMRILKETEL 720  
Dh 661 ILLVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOQMRILKETEL 720  
Qy 721 RKVKVLSGAFGVVYKGIWIPGGENYKIVAIKVRBNTSPKANKETILDEAYYMAVGSP 780  
Dh 721 RKVKVLSGAFGVVYKGIWIPGGENYKIVAIKVRBNTSPKANKETILDEAYYMAVGSP 780  
Qy 781 YYSRLIGICLTSTVOLVQLMPYGLLDHVRNRRGLSGODLLNMCQIAKGSYLEDYR 840  
Dh 781 YYSRLIGICLTSTVOLVQLMPYGLLDHVRNRRGLSGODLLNMCQIAKGSYLEDYR 840  
Qy 841 LVHRDLAANVLYKSPBNHKTIDFGIARLLIDETRYHADGKVPKMMALLESILRRFT 900  
Dh 841 LVHRDLAANVLYKSPBNHKTIDFGIARLLIDETRYHADGKVPKMMALLESILRRFT 900  
Qy 901 HOSDWSYGVVWELMTFAKRYDGI PAEIPDLEKGRLLPQPCITIDVYIMYKCM 960  
Dh 901 HOSDWSYGVVWELMTFAKRYDGI PAEIPDLEKGRLLPQPCITIDVYIMYKCM 960  
Qy 961 IDSECRPRRELVSERSMARDPORFVYQNEIDLGPASPLDSTFYSLLEDDMGDLVDA 1020  
Dh 961 IDSECRPRRELVSERSMARDPORFVYQNEIDLGPASPLDSTFYSLLEDDMGDLVDA 1020  
Qy 1021 EEYLVPQOGFCPCDPAPGAGVHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080  
Dh 1021 EEYLVPQOGFCPCDPAPGAGVHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080  
Qy 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
Dh 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
Qy 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200

Db 1141 NOPDVRPDPSPREBPLPAPARPAATLEPKTLSPKNGVVDVAFGAVNPEVLPQ 1200  
Qy 1201 GGAAPQHPHPAPFAPFADNLYYMDODPEERGAPESTFKGTPTAENPEVYGLDVPV 1255  
Db 1201 GGAAPQHPHPAPFAPFADNLYYMDODPEERGAPESTFKGTPTAENPEVYGLDVPV 1255  
RESULT 12  
AAE20479  
ID AAE20479 standard; Protein; 1255 AA.  
XX AAE20479;  
XX AAE20479;  
XX 01-JUL-2002 (first entry)  
XX Human Her-2/neu protein.  
XX Human Her-2/neu protein.  
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;  
XX human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Region 1021..1030  
XX /note="Naturally processed HLA-B44-restricted epitope"  
XX WO200214503-A2.  
XX 21-FEB-2002.  
XX 14-AUG-2001; 2001WO-US41733.  
XX 14-AUG-2000; 2000US-225152P.  
XX 28-SEP-2000; 2000US-236428P.  
XX 21-FEB-2001; 2001US-270520P.  
XX (CORI-) CORIYA CORP.  
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodee MJ, Kalos MD;  
XX Mcneill PD, Vedvick TS;  
XX MPI: 2002-280758/32.  
XX N-PSDB; AAD32743.  
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
XX prevention and diagnosis of cancer, preferably breast cancer  
XX Disclosure; Page 114-117; 129pp; English.  
XX The invention relates to an isolated Her-2/Neu polypeptide composition  
XX effective for eliciting an immune response. The invention is useful for  
XX eliciting an immune response in a patient, where the patient is human  
XX leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.  
XX The composition is useful for the therapy and diagnosis of cancer,  
XX preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
XX and other compositions for the diagnosis, prevention and treatment of  
XX human malignancies, for stimulating and/or expanding T cells specific for  
XX Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
XX patient. The invention is useful for stimulating a T cell response in a  
XX human patient, as probe or primer for nucleic acid hybridisation, to  
XX selectively form duplex molecules with complementary stretches of the  
XX entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
XX length gene from a suitable library, and to direct expression of a  
XX polypeptide in appropriate host cells. The composition is useful in  
XX prophylactic or therapeutic applications and for the treatment of cancer,  
XX preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
XX associated malignancies. The invention is useful in gene therapy. The  
XX present sequence is human Her-2/neu protein.  
SQ Sequence 1255 AA;  
Query Match 98.9%, Score 6734; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
Qy 1 METALACMGILLALLPGGAASVTOVCTGDMTLRLPASETHLMDLRHLVOCGVVQGNL 60  
Db 1 METALACMGILLALLPGGAASVTOVCTGDMTLRLPASETHLMDLRHLVOCGVVQGNL 60  
Qy 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVROVPLQRLIRVGTOLFEDNALAVLDNG 120  
Db 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVROVPLQRLIRVGTOLFEDNALAVLDNG 120  
Qy 121 DPLNNTTPTVGTASPGGLRELOLRSLTEILKGVLIQRNPOLCYDITLWKDXYIKANSKF 180  
Db 121 DPLNNTTPTVGTASPGGLRELOLRSLTEILKGVLIQRNPOLCYDITLWKDXYIKANSKF 180  
Qy 181 IGITELNRSRACHPCSPCKGSRCKGSESDQSILTRVCAGGACRCGRLPTDCCHECC 240  
Db 181 LTLIDTNSRACHPCSPCKGSRCKGSESDQSILTRVCAGGACRCGRLPTDCCHECC 240  
Qy 241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTTFGASCVTACP 300  
Db 241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTADGTORCEKSKPCARVCYGLMEHLREVRVTSAN 360  
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTADGTORCEKSKPCARVCYGLMEHLREVRVTSAN 360  
Qy 361 IOEPAGCKKIRGSLAFLESEFDGPASNTAPQAPOLVFEFLBEITGLYISAMPDLSLP 420  
Db 361 IOEPAGCKKIRGSLAFLESEFDGPASNTAPQAPOLVFEFLBEITGLYISAMPDLSLP 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGGISWLGRLSRLRELGSLALIHNNHLCFVHV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGGISWLGRLSRLRELGSLALIHNNHLCFVHV 480  
Qy 481 PMDOLFRNPQALLHTANRPEDECVEGLACHOLCARGCMGPGPTCCVNCQSFRLRGDEC 540  
Db 481 PMDOLFRNPQALLHTANRPEDECVEGLACHOLCARGCMGPGPTCCVNCQSFRLRGDEC 540  
Qy 541 VEECRVLOGLPREVYNARHCIPCHPECCPONGSVTCRFPBADOCVACAHYKDPFCVARC 600  
Db 541 VEECRVLOGLPREVYNARHCIPCHPECCPONGSVTCRFPBADOCVACAHYKDPFCVARC 600  
Qy 601 PSGVCPDLSYMPIKWFPEDEGACOPCPINCHSCVYDLDDKCPAQRASPLTISAVVG 660  
Db 601 PSGVCPDLSYMPIKWFPEDEGACOPCPINCHSCVYDLDDKCPAQRASPLTISAVVG 660  
Qy 661 ILLVVVLGVVEGILIKRQOKIRRYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVEGILIKRQOKIRRYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVYVLGSGAGTYKGIWIDGENVKIPVAKYLRRENTSPRANKELIDEAYVWAGVSP 780  
Db 721 RKVYVLGSGAGTYKGIWIDGENVKIPVAKYLRRENTSPRANKELIDEAYVWAGVSP 780  
Qy 781 YVSRLLGICLSTYQVLTQMLPYGCLDHWENRGLSGDQLMCMQIAGMSVLEVR 840  
Db 781 YVSRLLGICLSTYQVLTQMLPYGCLDHWENRGLSGDQLMCMQIAGMSVLEVR 840  
Qy 841 LVHBDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900  
Db 841 LVHBDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900  
Qy 901 HQSVNVSYGTVWELMTFGAKPYGIPARELPDLLEGERLPQPPICITIVYIMVCKM 960  
Db 901 HQSVNVSYGTVWELMTFGAKPYGIPARELPDLLEGERLPQPPICITIVYIMVCKM 960  
Qy 961 IDSECRFPRELVESEFRMARDDPORFVYIWNEDIGPASPLDSTYRSLLEDDMDGLVDA 1020  
Db 961 IDSECRFPRELVESEFRMARDDPORFVYIWNEDIGPASPLDSTYRSLLEDDMDGLVDA 1020  
Qy 1021 EBYLVPOQGFCDPAPAGGMVHRHRSSTSGGDLTLGLPSESEAPRSLPASEG 1080  
Db 1021 EBYLVPOQGFCDPAPAGGMVHRHRSSTSGGDLTLGLPSESEAPRSLPASEG 1080

Db 1021 EBYLVPOQGFCCPDPAAGAGVWHRHRSSSTRSGCGDLTLGLPSEBEPSPPLAPSEG 1080  
QY 1081 AGSDVFPDDLGGAAGKGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVALTCSPQPEYV 1140  
Db 1081 AGSDVFPDDLGGAAGKGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVALTCSPQPEYV 1140  
QY 1141 NOPDVRPOPSPREBPPLPAARPAAGATLERAKTSLPGKNGVADVAFGAVENPEYLTPO 1200  
Db 1141 NOPDVRPOPSPREBPPLPAARPAAGATLERAKTSLPGKNGVADVAFGAVENPEYLTPO 1200  
QY 1201 GGAAPOHPHPAFSPAFDNLVYWDODPPERGAAPSTFKCTPTAENPEYLGIDVPV 1255  
Db 1201 GGAAPOHPHPAFSPAFDNLVYWDODPPERGAAPSTFKCTPTAENPEYLGIDVPV 1255

## RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;

DT 17-JUN-2002 (first entry)

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

KM Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

KM tyrosine kinase; receptor; c-erbB2; gene therapy.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Domain 1..653 /note="extracellular domain"

FT Domain 676..1255 /note="intracellular domain"

FT Domain 990..1255 /note="phosphorylation domain"

PN WO200212341-A2.

PD 14-FEB-2002.

PF 03-AUG-2001; 2001WO-US24283.

PR 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Cheever MA, Gheysen D;

DR WPI: 2002-241743/29.

DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

PT or enhancing an immune response to the protein, has Her-2/neu

PT extracellular domain fused to Her-2/neu intracellular or

PT phosphorylation domain -

PS Claim 68; Fig 7; 14pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein  
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer  
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed  
CC in a variety of cancers, including breast, ovarian, colon, lung and  
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase  
CC family of receptor-like glycoproteins. It comprises an extracellular  
CC domain with homology to the epidermal growth factor receptor  
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal  
CC intracellular domain that also shows homology to EGFR. Its  
CC overexpression correlates with a poor prognosis in breast and  
CC ovarian cancers. The invention provides Her-2/neu fusion  
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In  
CC preferred fusion proteins, the extracellular domain of a Her-2/neu  
CC protein is fused to a Her-2/neu intracellular domain or  
CC phosphorylation domain (or its Deltap fragment). An immune  
CC response to Her-2/neu protein is elicited or enhanced by  
CC administering the fusion protein in the form of a vaccine, or by  
CC transfecting cells of an animal *ex vivo* with a nucleic acid  
CC encoding the fusion protein, and delivering the transfected cells  
CC to the animal. The fusion proteins, nucleic acids, and isolated  
CC specific T-cells are useful for inhibiting the development of a  
CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
CC in a patient. T cells that specifically react with a Her-2/neu  
CC fusion protein can be used to remove tumour cells from a sample in  
CC order to inhibit the development of cancer in a patient.

SQ Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MELALCWGILLALLPFGAASVYCTGTDKMLRLPASBETHLMLRHLVYGCQVVGSL 60  
Db 1 MELALCWGILLALLPFGAASVYCTGTDKMLRLPASBETHLMLRHLVYGCQVVGSL 60  
QY 61 ETTYPTNASISFLDIOEVQGYVLIANNOVQVLOLRIRVGTQLFEDNALAVLDNG 120  
Db 61 ETTYPTNASISFLDIOEVQGYVLIANNOVQVLOLRIRVGTQLFEDNALAVLDNG 120  
QY 121 DPLNNTTPTVTSAPGGLRELOLRSLTEILKGGVLIQRNPOLCYDPTILMKDQYIRANSKF 180  
Db 121 DPLNNTTPTVTSAPGGLRELOLRSLTEILKGGVLIQRNPOLCYDPTILMKDQYIRANSKF 180  
QY 121 DPLNNTTPTVTSAPGGLRELOLRSLTEILKGGVLIQRNPOLCYDPTILMKDQYIRANSKF 180  
Db 121 DPLNNTTPTVTSAPGGLRELOLRSLTEILKGGVLIQRNPOLCYDPTILMKDQYIRANSKF 180  
QY 181 IGITELANSRACHPESPMCKSKRCWESSEDCQSLTRYVACAGCARCKPLPTDCHEQC 240  
Db 181 IGITELANSRACHPESPMCKSKRCWESSEDCQSLTRYVACAGCARCKPLPTDCHEQC 240  
QY 241 AAGCTGPRHSDCLALHNHSGICELHCPALVYTTDTFESMPNEGRTTFASCVTACP 300  
Db 241 AAGCTGPRHSDCLALHNHSGICELHCPALVYTTDTFESMPNEGRTTFASCVTACP 300  
QY 301 YNYLSTDVGSCTLYVCPPLNNOEVTAEADGTORCEKSKPCARVCYGLMEHLREVRVTSAN 360  
Db 301 YNYLSTDVGSCTLYVCPPLNNOEVTAEADGTORCEKSKPCARVCYGLMEHLREVRVTSAN 360  
QY 361 IOEPAGCKKIRGSLAFLEPSPDGDPASNTAPLPQPOLQVFETLEITGLYLSAMPDSL 420  
Db 361 IOEPAGCKKIRGSLAFLEPSPDGDPASNTAPLPQPOLQVFETLEITGLYLSAMPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNHGAVALTLQGLISMLGLRSLRELGSGLALHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNHGAVALTLQGLISMLGLRSLRELGSGLALHNNHLCFVHTV 480  
QY 481 PWDOLFRRPHQALLHTANRPBDECVBEGGLACHQLCARHGWGPGTQCVNCSOFLRGQEC 540  
Db 481 PWDOLFRRPHQALLHTANRPBDECVBEGGLACHQLCARHGWGPGTQCVNCSOFLRGQEC 540  
QY 541 VEECRVLDGLPREYVNAHCLPCHRECQPNQSVTCFGBEADQCAACHYKDPPECVARC 600  
Db 541 VEECRVLDGLPREYVNAHCLPCHRECQPNQSVTCFGBEADQCAACHYKDPPECVARC 600  
QY 601 PSGVPRDLSYMPIMKFPBEGACQPCPINCCHSCVLDLDDKCPAQRASPLTSISAVVG 660  
Db 601 PSGVPRDLSYMPIMKFPBEGACQPCPINCCHSCVLDLDDKCPAQRASPLTSISAVVG 660  
QY 661 ILVVVILGVFEGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAOQRILKETEL 720  
Db 661 ILVVVILGVFEGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAOQRILKETEL 720  
QY 721 RKVKVLSGAGCTVYKGIWIDGENVKIPVAIKVLRNTSPRANKETIDEAVMAGVQSP 780  
Db 721 RKVKVLSGAGCTVYKGIWIDGENVKIPVAIKVLRNTSPRANKETIDEAVMAGVQSP 780

```

QY 781 YVSRLLGICLTSTVOLVTOLMPYGCILDHVRNENRGLSGSODLLNMCMQIAKMSYLEDR 840
DB 781 YVSRLLGICLTSTVOLVTOLMPYGCILDHVRNENRGLSGSODLLNMCMQIAKMSYLEDR 840
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHYADGGKVPILKMALESILRRRT 900
DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHYADGGKVPILKMALESILRRRT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMIMVKCM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMIMVKCM 960
QY 961 IDSECRPRRELVSERSRMARDPQRFVVIQNEIDLGPASPLDSTFYSLLEDMDGLVDA 1020
DB 961 IDSECRPRRELVSERSRMARDPQRFVVIQNEIDLGPASPLDSTFYSLLEDMDGLVDA 1020
QY 1021 EBYLVPOOGFCPPDAPAGAGVWHRHRSSTRSGGDDTLGLEPSEEEAPRSPILASRG 1080
DB 1021 EBYLVPOOGFCPPDAPAGAGVWHRHRSSTRSGGDDTLGLEPSEEEAPRSPILASRG 1080
QY 1081 AGSDVPDGDLGWAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVPDGDLGWAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPSRREGPLPAARAGATLEBAKTLSEKGVKQVQVAFGAVENPEYLPQ 1200
DB 1141 NOPDVRPQPSRREGPLPAARAGATLEBAKTLSEKGVKQVQVAFGAVENPEYLPQ 1200
QY 1201 GGAAPQHPPPAFSAFDNLYYMDOPPERGAPPSSTFKGTPTAENPEYLGIDVPV 1255
DB 1201 GGAAPQHPPPAFSAFDNLYYMDOPPERGAPPSSTFKGTPTAENPEYLGIDVPV 1255

```

RESULT 14  
AAU77114 standard; Protein; 1255 AA.

```

ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human, Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN W0200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
XX
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
DR WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

```

```

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA:
Query Match 98.9%; Score 6734; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 METALCRWGLLALLPGAASTOVCTGDMKLRPASETHLDMRLHYOGCQVQGNL 60
DB 1 METALCRWGLLALLPGAASTOVCTGDMKLRPASETHLDMRLHYOGCQVQGNL 60
QY 61 ELTYLPTNASLFLDIDIEVOGYVLIANQVQVPLQRLIRYRGTQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLFLDIDIEVOGYVLIANQVQVPLQRLIRYRGTQLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQLRSLEILKGGVLIQRNPOLCYODTILWKDQYIKANSXF 180
DB 121 DPLNNTTPTVGTASPGGLRELQLRSLEILKGGVLIQRNPOLCYODTILWKDQYIKANSXF 180
QY 181 IGITELNRSRACHPCSPCKSGRCSSESDQSLTRTVACAGCARCKGRLPTDCHECQ 240
DB 181 LTLIDTNRSRACHPCSPCKSGRCSSESDQSLTRTVACAGCARCKGRLPTDCHECQ 240
QY 241 AAGCTGPRHSDCLAGHNRHSGICELHCPALVYNTDTFESHPNREGRTTFCASCYTACP 300
DB 241 AAGCTGPRHSDCLAGHNRHSGICELHCPALVYNTDTFESHPNREGRTTFCASCYTACP 300
QY 301 YNYLSTDVGSCTVCPPLHNOEYTAEDGTORCEKSKPCARVCYGGMEHLRVRVAVTSAN 360
DB 301 YNYLSTDVGSCTVCPPLHNOEYTAEDGTORCEKSKPCARVCYGGMEHLRVRVAVTSAN 360
QY 361 IOEFAGCKKIFGSLAFLEPESFDGPASNTAPLOPBOLQVFBTEITGLYISAMPDSL 420
DB 361 IOEFAGCKKIFGSLAFLEPESFDGPASNTAPLOPBOLQVFBTEITGLYISAMPDSL 420
QY 421 DLSPFQNLQVIRGRILHNGAVSLTLQGLISWLGILRSIRELSSGIALIHHNTHLCFYHTV 480
DB 421 DLSPFQNLQVIRGRILHNGAVSLTLQGLISWLGILRSIRELSSGIALIHHNTHLCFYHTV 480
QY 481 PMDQLFRNPQALHTANRPEDECYGEGLACHOLCARGHCGPPTOCVNCQSQFLRGQEC 540
DB 481 PMDQLFRNPQALHTANRPEDECYGEGLACHOLCARGHCGPPTOCVNCQSQFLRGQEC 540
QY 541 VEECRVILQGLREYVNAHCLPCHPECCOPONGSVTCFPEADQCVACAHYDPPCVARC 600
DB 541 VEECRVILQGLREYVNAHCLPCHPECCOPONGSVTCFPEADQCVACAHYDPPCVARC 600
QY 601 PSQVPLDSTYPIWKFPDEBGACOPCPINCHSCVDLDDKCCPAQORASPLTISIAYVG 660
DB 601 PSQVPLDSTYPIWKFPDEBGACOPCPINCHSCVDLDDKCCPAQORASPLTISIAYVG 660
QY 661 ILVVVVLGVVFGILIKRQOKIRKYTRRLQETELVEPLTPSGAMPQAOAMRLIKETEL 720
DB 661 ILVVVVLGVVFGILIKRQOKIRKYTRRLQETELVEPLTPSGAMPQAOAMRLIKETEL 720
QY 721 RKVVVLSSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVWAGVSP 780
DB 721 RKVVVLSSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVWAGVSP 780
QY 781 YVSRLLGICLTSTVOLVTOLMPYGCILDHVRNENRGLSGSODLLNMCMQIAKMSYLEDR 840
DB 781 YVSRLLGICLTSTVOLVTOLMPYGCILDHVRNENRGLSGSODLLNMCMQIAKMSYLEDR 840
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHYADGGKVPILKMALESILRRRT 900

```

DB 841 LVHRDLAARNVLVKSFNHVKITDFGLARLDDIDETFEYHADGSKVPIKMMALESILRRRT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960  
QY 961 IDSECRPRPRELVSEFSRMAPRPFVYVQNEIDLGPASPLDSTFYSRLLEDMDMDGLVNA 1020  
DB 961 IDSECRPRPRELVSEFSRMAPRPFVYVQNEIDLGPASPLDSTFYSRLLEDMDMDGLVNA 1020  
QY 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEBEAPRSLAPSEG 1080  
DB 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEBEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAKGLQSLPETHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAKGLQSLPETHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1141 NOPDVRPOPSPREGPLAPAPAGATLEBAKTLSPGKNGVWVDVAFGGAIVENPEYLTPO 1200  
DB 1141 NOPDVRPOPSPREGPLAPAPAGATLEBAKTLSPGKNGVWVDVAFGGAIVENPEYLTPO 1200  
QY 1201 GGAAPQPHPPAPFSPAFDNLVYWDODPPERGA PPTSTFKGTPTAENPEYLGIDVPV 1255  
DB 1201 GGAAPQPHPPAPFSPAFDNLVYWDODPPERGA PPTSTFKGTPTAENPEYLGIDVPV 1255  
RESULT 15  
AAR39568  
ID AAR39568 standard; Protein; 1433 AA.  
XX AAR39568;  
XX  
XX 07-FEB-1994 (first entry)  
XX  
XX Sequence of c-erbB-2 tumour antigen.  
XX  
XX Tumour antigen: c-erbB-2; glycoprotein.  
XX  
XX Homo sapiens.  
XX  
XX WO9316185-A.  
XX  
XX PD 19-AUG-1993.  
XX  
XX PF 05-FEB-1993; 93WO-US01055.  
XX  
XX PR 06-FEB-1992; 92US-0831967.  
XX  
XX PA (CEFTU ) CEFTUS ONCOLOGY CORP.  
XX (CREA-) CREATIVE BIOMOLECULES INC.  
XX  
XX Houston LT, Huston JS, Oppermann H, Ring DB;  
XX  
XX DR WPI: 1993-272889/34.  
XX N-PSDB; AAQ46083.  
XX  
XX PT New single chain Fv polypeptide binding to C-erbB-2 tumour  
XX antigen - for imaging or treating breast or ovarian cancer etc.  
XX  
XX PS Disclosure; pages 48-54; 87pp; English.  
XX  
XX CC c-erbB-2 refers to a protein antigen expressed on the surface of  
XX tumour cells, such as breast and ovarian tumour cells, which is an  
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric  
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents  
XX the location of a stop codon in AAQ46083.  
XX  
XX Sequence 1433 AA;

Query Match 98.3%; Score 6691; DB 14; Length 1433;  
Best Local Similarity 98.2%; Pred. No. 0;

Matches 1233; Conservative 7; Mismatches 15; Indels 0; Gaps 0;  
QY 1 MELALCKWGLLALLPFGAASVCTGTDMKALRPASPEHLDMLRHLVQSCVQVQNL 60  
DB 1 MELALCKWGLLALLPFGAASVCTGTDMKALRPASPEHLDMLRHLVQSCVQVQNL 60  
QY 61 ELTYLPTNALSFLDIOEVQGYVLIANQVQVLPQRLRIYRGTLQFEDNALAVLNDG 120  
DB 61 ELTYLPTNALSFLDIOEVQGYVLIANQVQVLPQRLRIYRGTLQFEDNALAVLNDG 120  
QY 121 PLANTTPVTGASPGELRELOLSLTELKGGVLIQRNPQLCYQDTILMKDQYIKANSKF 180  
DB 121 PLANTTPVTGASPGELRELOLSLTELKGGVLIQRNPQLCYQDTILMKDQYIKANSKF 180  
QY 181 IGITELNRSRACHPSCSPCKSGRCGSESEDCOSLTRTYCAGGACRCKPRLTDCHEOC 240  
DB 181 IGITELNRSRACHPSCSPCKSGRCGSESEDCOSLTRTYCAGGACRCKPRLTDCHEOC 240  
QY 241 AAGCTGPRKHSDDLACLFHNSGICELHCPALVTYNTDPESPNPEGRTTFGASCVTACP 300  
DB 241 AAGCTGPRKHSDDLACLFHNSGICELHCPALVTYNTDPESPNPEGRTTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPPLHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRATVSAN 360  
DB 301 YNYLSTDVGSCTLVCPPLHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRATVSAN 360  
QY 361 IOEPACCKKIFGSLAPFLBESFGDPASNTAPLOPEQLOVFEETLEITGYLYISAMPDSL 420  
DB 361 IOEPACCKKIFGSLAPFLBESFGDPASNTAPLOPEQLOVFEETLEITGYLYISAMPDSL 420  
QY 421 DLSVFQNLQVLRGRILHNQAVSLTLQGLISWLGRLSRLBEGSGLALIHNNHLCGVHIV 480  
DB 421 DLSVFQNLQVLRGRILHNQAVSLTLQGLISWLGRLSRLBEGSGLALIHNNHLCGVHIV 480  
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGIALCHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGIALCHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
QY 541 VEECRVLOGLPREYVNAHCLPCHPEBCOPONGSVTCFGEADQCAACAHYKDPFCVARC 600  
DB 541 VEECRVLOGLPREYVNAHCLPCHPEBCOPONGSVTCFGEADQCAACAHYKDPFCVARC 600  
QY 601 PSGVPRDLSYMPIMKFPDEBEGACOPCINCHSCVDLDKGPABEORASPLTSYSAVVG 660  
DB 601 PSGVPRDLSYMPIMKFPDEBEGACOPCINCHSCVDLDKGPABEORASPLTSYSAVVG 660  
QY 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLOETELVEBLTPSGAMPNQAQRILKETEL 720  
DB 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLOETELVEBLTPSGAMPNQAQRILKETEL 720  
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILDEAVYMAVGSP 780  
DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILDEAVYMAVGSP 780  
QY 781 YVSRLLGICLTSTVOLVQMLMPYGLLDHVRNBRGLSODLLNMQOIAKMSYLEDVR 840  
DB 781 YVSRLLGICLTSTVOLVQMLMPYGLLDHVRNBRGLSODLLNMQOIAKMSYLEDVR 840  
QY 841 LVHRDLAARNVLVKSFNHVKITDFGLARLDDIDETFEYHADGSKVPIKMMALESILRRRT 900  
DB 841 LVHRDLAARNVLVKSFNHVKITDFGLARLDDIDETFEYHADGSKVPIKMMALESILRRRT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960  
QY 961 IDSECRPRPRELVSEFSRMAPRPFVYVQNEIDLGPASPLDSTFYSRLLEDMDMDGLVNA 1020  
DB 961 IDSECRPRPRELVSEFSRMAPRPFVYVQNEIDLGPASPLDSTFYSRLLEDMDMDGLVNA 1020  
QY 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEBEAPRSLAPSEG 1080  
DB 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEBEAPRSLAPSEG 1080

QY 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1141 NOPVVRPQPPSPREGPLPAARPAAGATLERAKTILSPCKNGVVDVFAFGAVENPEYLTPO 1200  
Db 1141 NOPVVRPQPPSPREGPLPAARPAAGATLERAKTILSPCKNGVVDVFAFGAVENPEYLTPO 1200  
QY 1201 GGAAPQPHPPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQPHPPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

Search completed: July 22, 2003, 08:40:39  
Job time : 43.9774 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds  
(without alignments)

5267.077 Million cell updates/sec

Title: SEQ4-210-224-12

Perfect score: 6789  
Sequence: 1 MELAALCRWGLLLALPPCA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6698	98.7	1255	1 A24571	protein-tyrosine k
2	5880	86.6	1260	1 TVRTNU	protein-tyrosine k
3	5876.5	86.6	1254	2 I48161	p-185 precursor -
4	3076	45.3	1210	1 GORUR	epidermal growth f
5	3052	45.0	1210	2 A53183	epidermal growth f
6	3031.5	44.7	1223	1 TVCHLV	epidermal growth f
7	2909.5	42.9	1308	2 A47253	epidermal growth f
8	2609	38.4	1166	1 S06142	protein-tyrosine k
9	2345.5	34.5	1342	2 A36223	kinase-related tra
10	2260.5	33.3	1339	2 J04387	epidermal growth f
11	1766.5	26.0	698	1 TVFVLV	epidermal growth f
12	1703	25.1	604	1 TVYUHL	protein-tyrosine k
13	1647	24.3	544	1 S35745	protein-tyrosine k
14	1640	24.2	545	2 S00727	protein-tyrosine k
15	1623	23.9	540	2 B44776	kinase-related tra
16	1621	23.9	540	1 TVFVEB	protein-tyrosine k
17	1569.5	23.1	1330	1 GQFBE	protein-tyrosine k
18	1444	21.3	644	2 A36325	epidermal growth f
19	1255	18.5	1323	2 E88257	epidermal growth f
20	1255	18.5	1374	2 S70712	protein-tyrosine k
21	1175	17.3	1369	2 S70713	protein-tyrosine k
22	1155	17.0	1717	1 A45558	epidermal growth f
23	1063	15.7	527	2 A42032	epidermal growth f
24	914.5	13.5	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 T43220	protein-tyrosine k
27	675	9.9	1363	1 INHUR	insulin-like growt
28	675	9.9	1382	1 INHUR	insulin receptor p
29	666	9.8	1383	2 A36080	insulin receptor p

30	665.5	9.8	1372	2 A34157	insulin receptor p
31	662	9.8	1607	2 T43212	insulin-like growt
32	658.5	9.7	1200	2 A36502	insulin receptor-r
33	647	9.5	1268	2 B36502	insulin receptor-r
34	642	9.5	1477	2 T18534	protein-tyrosine k
35	603	8.9	1367	1 IGHUR1	insulin-like growt
36	595	8.8	987	2 A54092	protein-tyrosine k
37	595	8.8	1371	2 A33837	insulin-like growt
38	593.5	8.7	2148	1 A56081	insulin receptor -
39	589.5	8.7	977	2 S49004	tyrosine kinase Mp
40	589	8.7	987	2 I48652	mouse developmenta
41	588	8.7	1114	1 S05582	protein-tyrosine k
42	588	8.7	1390	2 T30346	insulin receptor -
43	588	8.7	2101	2 S57245	insulin receptor (
44	586	8.6	1091	2 S33596	protein-tyrosine k
45	584.5	8.6	976	2 A36355	protein-tyrosine k

#### ALIGNMENTS

##### RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N.Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein e  
C.Species: Homo sapiens (man)  
C.Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #ext change 11-Jun-1999

C.Accession: A24571; A25491; A44188; B44188; I59509; I57622

R.Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T  
Nature 319, 230-234, 1986

A.Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth  
A.Reference number: A24571; MUID:86118663; PMID:3003577

A.Accession: A24571  
A.Molecule type: mRNA

A.Residues: 1-1255 <YAM>  
A.Cross-references: GB:X03363; NID:931197; PIDN:CAA27060.1; PID:931198

R.Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A.Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid  
A.Reference number: A25491; MUID:86016729; PMID:2995967

A.Accession: A25491  
A.Molecule type: DNA

A.Residues: 737-1031 <SEM>  
A.Cross-references: GB:M1167; NID:9182163; PIDN:AAA35808.1; PID:9553282

R.Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg  
Science 230, 1132-1139, 1985

A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro  
A.Reference number: A44188; MUID:86070181; PMID:2999974

A.Accession: A44188  
A.Molecule type: DNA

A.Residues: 740-910 <COU>  
A.Cross-references: GB:M12036; NID:9183988; PIDN:AAA35978.1; PID:9183989

A.Accession: B44188  
A.Molecule type: mRNA

A.Residues: 1-517, 'RBL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU>  
A.Cross-references: GB:M11730; NID:9183986

R.King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985

A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A.Reference number: I59509; MUID:85272597; PMID:2992089

A.Accession: I59509  
A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA  
A.Residues: 832-909 <REX>

A.Cross-references: GB:L29395; NID:9459807; PIDN:AAA35809.1; PID:9459808  
R.Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987

A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcription  
A.Reference number: I57622; MUID:87286698; PMID:3033351

A.Accession: I57622  
A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA  
A.Residues: 1-191 <TAL>



A:Cross-references: GB:M6792; NID:9183983; PIDN:AAA58637.1; PID:9553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERBB2; NCL; NEU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 inase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-125/Domain: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:22-653/Domain: extracellular #status predicted <EXT>  
 F:70-304/Domain: EGF receptor extracellular domain repeat <BE1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <BE2>  
 F:654-675/Domain: transmembrane #status predicted <TMM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:68/124,187,253,571,629/Binding site: carboxylate (Asn) (covalent) #status predict  
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:753/Active site: Lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.7%; Score 6698; DB 1; Length 1255;  
 Best Local Similarity 98.8%; Pred. No. 4,6e-279;  
 Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

```

Oy 1 MELALCRWGLLLALPPGAATVOCTGDMKRLPASPEHLDMLRHLYGCGVYQGNL 60
Db 1 MELALCRWGLLLALPPGAATVOCTGDMKRLPASPEHLDMLRHLYGCGVYQGNL 60
Oy 61 ELTYLPTNASLSFLDIOEVQGVLIANHQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLDIOEVQGVLIANHQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
Oy 121 DPLNNTTPTVGTASPGGLRELORLTEILKGGVLIQORNPOLCYOPIILMKDIFHNQOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELORLTEILKGGVLIQORNPOLCYOPIILMKDIFHNQOLA 180
Oy 181 LTLIDTNSRACHPCSPMKSGRCSGSESDCCSLTRTVACAGCARCKPQYIKANSK 240
Db 181 LTLIDTNSRACHPCSPMKSGRCSGSESDCCSLTRTVACAGCARCKPQYIKANSK 240
Oy 241 FICITELKSDCLACLFHNSGICELHCALVTYNTDFESMPNBPGRYTFGASCTYACP 300
Db 241 FICITELKSDCLACLFHNSGICELHCALVTYNTDFESMPNBPGRYTFGASCTYACP 300
Oy 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSPCARVYGIAMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSPCARVYGIAMEHLREVRVTSAN 360
Oy 361 IOEPAGCKKIFGSLAFLESFDDGDPASNTAPLOPELOQVETLEITGLYISAMPDLSL 420
Db 361 IOEPAGCKKIFGSLAFLESFDDGDPASNTAPLOPELOQVETLEITGLYISAMPDLSL 420
Oy 421 DLSVQNIQVIRGRILHNGAVSLTIOGLISWLGIRSLRELSGIALIHNHNLCPVHTV 480
Db 421 DLSVQNIQVIRGRILHNGAVSLTIOGLISWLGIRSLRELSGIALIHNHNLCPVHTV 480
Oy 481 PMDOFRPHQALHTANRPEDECGEGGLAQOLCARHGWCPGPTQCVNCSQFIRGDEC 540
Db 481 PMDOFRPHQALHTANRPEDECGEGGLAQOLCARHGWCPGPTQCVNCSQFIRGDEC 540
Oy 541 VEECRVILQGLPREYVNAHCLPCHPECOFONGSVTCFGEADQCVAAHYKDPCEVARC 600
Db 541 VEECRVILQGLPREYVNAHCLPCHPECOFONGSVTCFGEADQCVAAHYKDPCEVARC 600
Oy 601 PSQVPRDLSYMPIMKFPDEEGACOPCPINCHSCVDLDDKGPAPQASPLTISIYAVVG 660
Db 601 PSQVPRDLSYMPIMKFPDEEGACOPCPINCHSCVDLDDKGPAPQASPLTISIYAVVG 660

```

```

Db 601 PSQVPRDLSYMPIMKFPDEEGACOPCPINCHSCVDLDDKGPAPQASPLTISIYAVVG 660
Oy 661 ILVVVLGVVFGILIKRROOKIRKTYMRLLOETELVEPLTPSGAMPNOAQRILKETEL 720
Db 661 ILVVVLGVVFGILIKRROOKIRKTYMRLLOETELVEPLTPSGAMPNOAQRILKETEL 720
Oy 721 RKVVLGSGAFGTVYKGIIPDGENVKIPVAIKVLRENTSPRANKIIDEAVMAGVGSF 780
Db 721 RKVVLGSGAFGTVYKGIIPDGENVKIPVAIKVLRENTSPRANKIIDEAVMAGVGSF 780
Oy 781 YVSRLLGICLTSTVOLVTOAMPYGLLDHVRNRRGLSQDILLNMQOIAKMSYLEVR 840
Db 781 YVSRLLGICLTSTVOLVTOAMPYGLLDHVRNRRGLSQDILLNMQOIAKMSYLEVR 840
Oy 841 LVHRLLAARNVYKSPNHYKLTDFGLARLIDIDENYHADGGKVPKMALESILRRRT 900
Db 841 LVHRLLAARNVYKSPNHYKLTDFGLARLIDIDENYHADGGKVPKMALESILRRRT 900
Oy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDVYMIWCKM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDVYMIWCKM 960
Oy 961 IDSECRPRRELVSFSSRMARDPQRFVVIQNEIDLCPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRRELVSFSSRMARDPQRFVVIQNEIDLCPASPLDSTFYRSLLEDMDGLVDA 1020
Oy 1021 EBYLVPOGCFPCPDPAAGCVMHRRSSSTRSGGDLTLGLESEEEAPRSPPLAPSG 1080
Db 1021 EBYLVPOGCFPCPDPAAGCVMHRRSSSTRSGGDLTLGLESEEEAPRSPPLAPSG 1080
Oy 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOKRSSEDPVPLPSETGCVAPPLCSQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOKRSSEDPVPLPSETGCVAPPLCSQPEYV 1140
Oy 1141 NQDVRPOPSPREBGPPLAARPAATLEBAKTLSPGKGVVDVFAFGAVENPEYLTPQ 1200
Db 1141 NQDVRPOPSPREBGPPLAARPAATLEBAKTLSPGKGVVDVFAFGAVENPEYLTPQ 1200
Oy 1201 GGAAPQPPPPAFAFNLNLYWDDPPERGAPESTFKTPAENPEYLGIDVPV 1255
Db 1201 GGAAPQPPPPAFAFNLNLYWDDPPERGAPESTFKTPAENPEYLGIDVPV 1255

```

## RESULT 2

TVRTNI

protein-tyrosine kinase (BC 2.7.1.112) neu precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1988 #sequence \_revision 31-Dec-1988 #text\_change 11-Jun-1999

C:Accession: A24562; A61204

R:Barzmann, C.I.; Hung, M.C.; Weinberg, R.A.

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein

A:Reference number: A24562; MUID:86118662; PMID:3945311

A:Accession: A24562

A:Molecule type: mRNA

A:Residues: 1-1260 &lt;BAR&gt;

A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746

R:Maatli, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen

Carcinogenesis 12, 1975-1978, 1991

A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals m

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A:Reference number: A61204; MUID:92035293; PMID:1682063

A:Accession: A61204

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 637-663, 'V', 665-702 &lt;MAS&gt;

A&gt;Note: authors translated the codon GCA for residue 25 as Val

C:Genetics: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

F:1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F:20-1260/Product: protein-tyrosine kinase neu #status predicted &lt;MAT&gt;

F:658-680/Domain: transmembrane #status predicted <TM>  
 F:723-988/Domain: protein kinase homology <KIN>  
 F:731-739/Region: protein kinase ATP-binding motif  
 F:71,191,263,535,576,634/Binding site: carbonyldiester (Aan) (covalent) #status predicted  
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:758/Active site: Lys #status predicted  
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.6%; Score 5880; DB 1; Length 1260;  
 Best Local Similarity 86.7%; Pred. No. 4.2e-244;  
 Matches 1090; Conservative 50; Mismatches 115; Indels 2; Gaps 2;

```

QY 1 MELAALCRWGLLALLPPGASTOVCTGDMKRLPASPETHLDMRLHYGCGVVOGNTL 60
DB 4 MELAAMCRWGLLALLPPGAGTGVCTGDMKRLPASPETHLDMRLHYGCGVVOGNTL 63
QY 61 ELTYLPTNASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDNYALAVLNDG 120
DB 64 ELTYVPANASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDNYALAVLNDR 123
QY 121 DPLNNTTTPVT-GASPGGLRELQSLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNQL 179
DB 124 DPONVAASTPGRTPEGLRELQSLRSLTEILKGVLIQRPOLCYODVWLMDVFRKNNQL 183
QY 180 ALTLIDNRSRACHPCSPMCKSGRCWGSSSDCSLTITVACGGACRCKPLPOYIKANS 239
DB 184 APVDIDNRSRACHPCACPKCKDNHCWGSPEDCQILITGTTISGACRCKGLPTDCCHEQ 243
QY 240 KFIITELKHSDDCLACHFNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCYTAC 299
DB 244 CAAGCTPRKSDCLACHFNHSGICELHCPALVTYNTDFESMPNBERRTFGASCYTTC 303
QY 300 PYNLTSDVGSCTLVCPLNQEVTAEDGTORCEKSKRCARVCYGLGMEHLREVAVTS 359
DB 304 PYNLTSTEVGSCTLVCPNQEVTAEADGTORCEKSKRCARVCYGLGMEHLRGARATISD 363
QY 360 NIOGFACCKTIFGSLAFLPESFDGDPASNTAPLOEQLQVETLEBETIGYLYISAMPDSL 419
DB 364 NIOGFACCKTIFGSLAFLPESFDGDPASNTAPLOEQLQVETLEBETIGYLYISAMPDSL 423
QY 420 PDLISFQNLQVIRGRILHNGAYSLTLOGLISWGLSLRLSGSLALIHNNHLCFPH 479
DB 424 RDLISFQNLQVIRGRILHNGAYSLTLOGLISWGLSLRLSGSLALIHNNHLCFPH 483
QY 480 VPMQDLFRNPHQALLHTANRPEDE-CYEGELACHQLCARGHCMWGPRTQVCNCSQPLRQ 538
DB 484 VPMQDLFRNPHQALLHTANRPEDE-CYEGELACHQLCARGHCMWGPRTQVCNCSQPLRQ 543
QY 539 ECVEECGVLOGLPREYVNAHNCLEPHECOPQNGSVTCFGEALQCVACAHYKORPCVA 598
DB 544 ECVEECGVLOGLPREYVNAHNCLEPHECOPQNGSVTCFGEALQCVACAHYKORPCVA 603
QY 599 RCEGVPAPDLISYMPKPFDEEGACOPCINCTHSCVDLDDKGPAPORASPLTISYAV 658
DB 604 RCEGVPAPDLISYMPKPFDEEGACOPCINCTHSCVDLDDKGPAPORASPLTISYAV 663
QY 659 VGLILVAVGVGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKET 718
DB 664 EGVLLFLILVAVGVGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKET 723
QY 719 ELAKVVLGSGAGCTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAVVMAVG 778
DB 724 ELAKVVLGSGAGCTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAVVMAVG 783
QY 779 SPYVSRLLGICLTSTVOLVQLMPYGLLDHVENRGRLSODPLNMCQIAGKMSYLED 838
DB 784 SPYVSRLLGICLTSTVOLVQLMPYGLLDHVENRGRLSODPLNMCQIAGKMSYLED 843
QY 839 VRLVHNDLAARNLVVSPNHVKITDFGLARLLDIDETEHADGGKVPDKMALESILRR 898
DB 844 VRLVHNDLAARNLVVSPNHVKITDFGLARLLDIDETEHADGGKVPDKMALESILRR 903
QY 899 FTHQSDVMSYGVTVVMEIMTFGAKRPYDGIIPAREIIPDLLEKGERLPQPICTIDVYIMVNC 958
  
```

```

DB 904 FTHQSDVMSYGVTVVMEIMTFGAKRPYDGIIPAREIIPDLLEKGERLPQPICTIDVYIMVNC 963
QY 959 WMIDSECKPRRELVESESRMAPRQRFVUQNEPLGASPLDSTFFYSLSLEDDMDGLV 1018
DB 964 WMIDSECKPRRELVESESRMAPRQRFVUQNEPLGASPLDSTFFYSLSLEDDMDGLV 1023
QY 1019 DAEETLVYQGGFFCCDPAPAGAGVHNRSSSTSGGDDTLGLPSESEAPRSPLAPS 1078
DB 1024 DAEETLVYQGGFFCCDPAPAGAGVHNRSSSTSGGDDTLGLPSESEAPRSPLAPS 1083
QY 1079 EGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOYSDEPTVPLPSETDGVAPLTCSPPE 1138
DB 1084 EGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOYSDEPTVPLPSETDGVAPLTCSPPE 1143
QY 1139 VYNOEDVPRPPSPREGVLPAAPRAGATLEBAKTLSFGKGVVMDVFAFGAVENPEYLT 1198
DB 1144 VYNOEDVPRPPSPREGVLPAAPRAGATLEBAKTLSFGKGVVMDVFAFGAVENPEYLT 1203
QY 1199 POGGAAPQPPPPAFSPAFDNLVYWDODPPRSGAPSTFFKGTPTAENPEYLGLDVVP 1255
DB 1204 PREGTASPRHSPAFSPAFDNLVYWDONSSGCGPPSPNFBGTPTAENPEYLGLDVVP 1260

```

RESULT 3  
 148161  
 P-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
 C:Accession: 148161  
 R:Nakamura, T.; Uchijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa  
 Gene 140, 251-255, 1994  
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Reference number: 148161; MUID:94193007; PMID:7908275  
 A:Accession: 148161  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <RES>  
 A:Cross-references: GB:D16295; NID:g493236; PID:BAA03801.1; PID:g747595  
 A:Genetic:  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.6%; Score 5876.5; DB 2; Length 1254;  
 Best Local Similarity 86.5%; Pred. No. 5.8e-244;  
 Matches 1086; Conservative 58; Mismatches 110; Indels 1; Gaps 1;

```

QY 1 MELAALCRWGLLALLPPGASTOVCTGDMKRLPASPETHLDMRLHYGCGVVOGNTL 60
DB 1 MELAAMCRWGLLALLPPGAGTGVCTGDMKRLPASPETHLDMRLHYGCGVVOGNTL 63
QY 61 ELTYLPTNASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDNYALAVLNDG 120
DB 64 ELTYVPANASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDNYALAVLNDR 123
QY 121 DPLNNTTTPVTGASPGGLRELQSLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNQL 180
DB 124 DPLDNVTATGRTPEGLRELQSLRSLTEILKGVLIQRPOLCYODVWLMDVFRKNNQL 183
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGSSSDCSLTITVACGGACRCKPLPOYIKANS 240
DB 184 LTLIDNRSRACHPCSPMCKSGRCWGSSSDCSLTITVACGGACRCKPLPOYIKANS 243
QY 241 FGIITELKHSDDCLACHFNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCYTAC 300
DB 244 AAGCTPRKSDCLACHFNHSGICELHCPALVTYNTDFESMPNBERRTFGASCYTTC 303
QY 301 PYNLTSDVGSCTLVCPLNQEVTAEDGTORCEKSKRCARVCYGLGMEHLREVAVTSAN 360
DB 304 PYNLTSTEVGSCTLVCPLNQEVTAEDGTORCEKSKRCARVCYGLGMEHLRGARATISAN 363
  
```

Qy	361	IOEFAGCKKIRGSLAFIPESDFDGPANSTAPLOPQLOVFPFLEETITGYLISAMPDLP	420
Dp	361	IOEPAGCKKIRGSLAFIPESDFDGNPSSGIALPTBPQLOVFETLEETITGYLISAMPDLSH	420
Qy	421	DLVSFQNLQVIRGRILNHGAYSLLTLOGIGISWLGSRSLRELSGSLALIHNNTHLCFVATV	480
Dp	421	DLVSFQNLQVIRGRVILHNGAYSLSLLOGIGIRWLGSRSLRELSGSLVLIHRNTHLCFVATV	480
Qy	481	PMDOLFRRPHQALLHTARPEDECYGBELACHOCARHCWGPGTQCVNCSQPLRGEC	540
Dp	481	PMDOLFRRPHQALLHSGNPSEEGCLKOPACPLCAHCHWGPGTQCVNCSHFIRGEC	540
Qy	541	VEECRVLOGLFREYVNAHCHLPCHEECOPONGSVYCFEPREADQVACAHYKDDPPECVARC	600
Dp	541	VKECRVMWKGILFREYVNGHCHLPCHEECQPONSTETCTGSEADQCACHYKDSPPCVARC	600
Qy	601	PSGVKPDLSYMPIMKCFPDEBEGACOPCPINCHSCYDLDDKCGPAEQRASPLTSIVSAVYG	660
Dp	601	PSGVKPDLSYMPIMKYCPDEBEGCOPCPINCHSCYDLDERGCPAQORASPATSIITATVYG	660
Qy	661	ILLVVLGVNVGILIKRROOKIRKTYMRLLQETLVEPLTPSGAMPQAOMRILKETEL	720
Dp	661	ILLPLVIGVNVGILIKRROROKIRKTYMRLLQETLVEPLTPSGAMPQAOMRILKETEL	720
Qy	721	RKVAVLSSGAFQTVYKGIWIPDGENVKIPVAIKVLENTSPKANKEILDEAYVMAVGSP	780
Dp	721	RKVAVLSSGAGTGYVKGIWIPDGENVKIPVAIKVLENTSPKANKEILDEAYVMAGLGSP	780
Qy	781	YVSRLGICLSTVOLVQOLMPYGGCLLDHVENRRGLSODLLNCOMOIAKMSVLEBYR	840
Dp	781	YVSRLGICLSTVOLVQOLMPYGGCLLDHVEHRRGLSODLLNCOMVQIAKMSVLEBYR	840
Qy	841	LVRHDLARNLVYKSPNHVKITDFGLARLLDIDETEHYADGKVPDKMMALESIIRRRFT	900
Dp	841	LVRHDLARNLVYKSPNHVKITDFGLARLLDIDETEHYADGSKVPDKIMIALESIRRRFT	900
Qy	901	HOSDVMSGCVVWELMTFGAKRPYDQIPAREIPDLLEKEERLPOPICITIDVYIMVKKMM	960
Dp	901	HOSDVMSGCVVWELMTFGAKRPYDQIPAREIPDLLEKEERLPOPICITIDVYIMVKKMM	960
Qy	961	IDSECRPFRELFVSEFSRMAARDPQRFVYIQNEIDLCPASPLDSTFASRLLEDODMDLYDA	1020
Dp	961	IDSECRPFRELFVSEFSRMAARDPQRFVYIQNEIDLGPSSPLDSTFASRLLEDODMDLYDA	1020
Qy	1021	EELYLPQOGFFCPDPAFAGAGMWHHRHSSSTRSGGDLTGLGEPSEEDAPRSLAPSEEG	1080
Dp	1021	EELYLPQOGFFCPDPAFAGAGTAHRHRHSSSTRSGGGLTGLGMEPSGEEPPRSLAPSEEG	1080
Qy	1081	AGSDVFDODLGMGAKKLOSLPTHNPSPFORSEDPYVPLPSETTGVAAPLTCSPQRPXYV	1140
Dp	1081	AGSDVFEELGMGAKKQPOSISRDLSPLOKXSEPTPLPLPETTGVAAPLACSPQRPXYV	1140
Qy	1141	NOPEVRPOPPLTEBEGPLPVPVRPAGATLTERPKLISGKKGVVADVTFGCAVENPEYLAVR	1200
Dp	1141	NOPEVRPOPPLTEBEGPLPVPVRPAGATLTERPKLISGKKGVVADVTFGCAVENPEYLAVR	1200
Qy	1201	GGAAPQHPAPAFSAFNDLYYMODDPERGAPSTFMTGTPAENPEYLGLDVPY	1255
Dp	1201	GGSASQHP-PPPLCDPFDNLTYMODDPERGSPPTFTEGTPAENPEYLGLDVPY	1254

RESULT 4  
 GENE  
 epidermal growth factor receptor precursor - human  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: Homo sapiens (man)  
 C>Date: 15-Nov-1994 Sequence revision 27-Nov-1995 #text change 11-Jun-1999  
 C:Accession: A00611, R25772; S30024, A38622; A00642, A43615; A23062, A05281; A60143; A333  
 R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y  
 19, P.H.  
 Nature 309, 418-425, 1984  
 Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312  
A:Accession: A00641  
A:Molecule type: mRNA  
A:Residues: 1-1210 <ULL>  
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA5240.1; PID:g757924  
A>Note: The authors translated the codon AAG for residue 540 as Asn  
R:Hall, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985  
A:Title: Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene  
A:Reference number: A25772; MUID:85270438; PMID:2991899  
A:Accession: A25772  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-29 <ISH>  
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA5370.1; PID:g553272  
R:Hailey, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.  
Oncogene Res. 1, 375-396, 1987  
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification of the coding sequence  
A:Reference number: S30024; MUID:88217333; PMID:3329716  
A:Accession: S30024  
A:Molecule type: DNA  
A:Residues: 1-29 <HA2>  
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119  
R:Hailey, J.D.; Waterfield, M.D.  
J. Biol. Chem. 266, 17461-1753, 1991  
A:Title: Contributory effects of de Novo transfection and premature transcript termination on the expression of the human epidermal growth factor receptor cDNA in a variety of RNR cell lines  
A:Reference number: A38672; MUID:91107677; PMID:1988448  
A:Accession: A38672  
A:Molecule type: DNA  
A:Residues: 1-29 <HAL>  
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA6171.1; PID:g553271  
A:Experimental source: carcinoma cell line A431-7  
R:Xu, Y.; Ishii, S.; Clark, A.J.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Hailey, J.D.  
Nature 309, 806-810, 1984  
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNR cell lines  
A:Reference number: A00642; MUID:84245835; PMID:6330553  
A:Accession: A00642  
A:Molecule type: mRNA  
A:Residues: 'RCMRA', 150-187, 'KSVQNAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-327, 'T', 329-354, 'T', 356-383, 'T', 385-411, 'R', 413-942 <XUY>  
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF receptors  
R:Lin, C.R.; Chen, W.S.; Krueger, W.; Scolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.M.  
Science 224, 843-848, 1984  
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification and characterization  
A:Reference number: A43615; MUID:84196372; PMID:6326251  
A:Accession: A43615  
A:Molecule type: mRNA  
A:Residues: 713-964 <LIN>  
A:Experimental source: epidermoid carcinoma cell line A431  
R:Stimmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.  
Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
A:Reference number: A23062; MUID:85046483; PMID:6093780  
A:Accession: A23062  
A:Molecule type: mRNA  
A:Residues: 1028-1210 <SIM>  
R:Weber, W.; Gyll, G.N.; Speiss, J.  
Science 224, 294-297, 1984  
A:Reference number: A05281; MUID:84172183; PMID:6324343  
A:Accession: A05281  
A:Molecule type: protein  
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>  
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.  
J. Biol. Chem. 260, 5205-5208, 1985  
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor  
A:Reference number: A60143; MUID:85182650; PMID:2985560  
A:Accession: A60143  
A:Molecule type: protein  
A:Residues: 740-744, 'X', 746-747 <RUB>  
R:McCoy, B.; Mosig, G.; Cohen, S.  
Nature 309, 270-273, 1984  
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase  
A:Reference number: A58023; MUID:84191554; PMID:6325948  
A:Contents: annotation; receptor activity

A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
R.Chen, W.S.; Lazar C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J  
Cell 59, 33-43, 1989  
A:Title: Functional independence of the epidermal growth factor receptor from a domain  
A:Reference number: A3331; MUID:90003233; PMID:2790960  
A:Contents: annotation; internalization signal  
C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c  
C:Genetics:  
A:Gene: GDB:EGFR  
A:Cross-references: GDB:120610; OMTM:131550  
A:Map position: 7p12.3-7p12.1  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-1210/Product: EGF receptor #status predicted <MT>  
F:25-645/Domain: extracellular #status predicted <EXT>  
F:75-300/Domain: EGF receptor extracellular domain repeat <E1>  
F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>  
F:646-668/Domain: transmembrane #status predicted <TM>  
F:669-1210/Domain: intracellular #status predicted <INT>  
F:710-975/Domain: protein kinase homology <KIN>  
F:718-726/Region: protein kinase ATP-binding motif  
F:999-1046/Region: coated-pit mediated internalization signal  
F:1047-1210/Region: inhibitory  
F:128,173,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predict  
F:745/Active site: Lys #status experimental

Query Match	45.3%	Score 3076	DB 1	Length 1210
Best Local Similarity	49.0%	Pred No. 2.5e-124		
Matches 620	Conservative 177	Mismatches 362	Indels 106	Gaps 21

[illegible]

Db	603	NNTL-VMKXADAGVCHLCHPNCTYGCCTGPGLEBGPNTGPKI P--STATGVMGALLLVLV	655
Qy	665	VVLGVVFGILLKRRQOKRKTTRRLLLOETVLVBPITSGAMPQAOQRILIKETELRVX	724
Db	660	VALGIG--LMRRRHIVRKTRTLRRLLOERLVEPLTPSGAPQALRIIRIKETEFKKIK	716
Qy	725	VLSGAFSTYKGLIIPGGENVKI PVAIKVLRENTSPKANEILIDEAVVMAGVSPYSR	784
Db	717	VLSGAFSTYKGLIIPGGEKVKI PVAIKELREATSPKANEILIDEAVVMASVDNPHYCR	776
Qy	785	LLGICLITSTVOLVTOLMRYGCLLDHVRENRGLCSODLLNCMQIANKMSYLEDVRLVHR	844
Db	777	LLGICLITSTVOLITOLMPFGCLLDVREHKKNISQOYLNNCVOIANKMNYLEBRRLVHR	836
Qy	845	DLAARNVLVKSPPNHVKTIDPGLAFLDIDETEVHADGKVIKMMALESTLRREFTHOSD	904
Db	837	DLAARNVLVKIPOHKVKTIDPGLAKLIGAEKEKYEHAEGKVPIMMALESLIHRITYTHOSD	896
Qy	905	VMSYGVTVWELMTFGAKRYDGI PARAEI PDLLEKEGRLPORPITCTIDVVMIMVKCMIMDE	964
Db	897	VMSYGVTVWELMTFGSKRPYDGI PASEISSLIEKEGRLPORPITCTIDVVMIMVKCMIMDAD	956
Qy	965	CRPRFRELVSFSFMAADPOAFVVIQ-NEDIGRASPILDSTFRSLLEDDMDGLVDABEY	1023
Db	957	SRPRFRELIIFFSFKADRPORYLVIQODERHNLSPDTSNFYRLMDEEDMDVDVADXEY	1016
Qy	1024	LVPOGFFPCRPAPGAGGMVHNHRSSSTRGCGDLTLGLEPSEEA PRSPAPSEGAS	1083
Db	1017	LIPQGGF-----SSPSTSRTPBLSSLSATS	1042
Qy	1084	DVEDGDJLGMGAKLOSLPTHDPSPLORYSEDPVLPSET--DGUYAPLTCSPQPEYVN	1141
Db	1043	N--NSTVACIDRNLOCSPIKEDESFLLGRYSDDPFGALTBDSDIDTFL-----PVPEYIN	1094
Qy	1142	QPDVRPQRPSPBGRSLPAARAGLTTERAKTLSFGKRGVYKVDYAFGCAVBNPEYL-TPQ	1200
Db	1095	Q-SYPKRPAAGVQVPHVNOPLN-----ABSRDPHYOD--PHSTAAGNPEYLLTVO	1143
Qy	1201	GGAAPQHPRPAPSPAFDNLYWMO-----DP-----PERGAPSTFGKGTAE	1244
Db	1144	-----PTCVNSTFDSRAHMQKSHQISLDNPNRYQODFFKXKACNIGFSGS-TAE	1193
Qy	1245	NPEYL 1249	
Db	1194	NAEYL 1198	

## RESULT 5

epidermal growth factor receptor precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 18-Jun-1999  
C:Accession: A53183; A43818; S24942; S45345; I49643  
R:Unetctee, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.  
Genes Dev. 8, 399-413, 1994  
A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor  
A:Reference number: A53183; MUID:94170986; PMID:812525  
A:Accession: A53183  
A:Molecule type: mRNA  
A:Residues: 1-1210 <LUE>  
A:Cross-references: GB:U03425  
R:Aviv, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
Oncogene 6, 673-676, 1991  
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding  
A:Reference number: A43818; MUID:91232866; PMID:2030916  
A:Accession: A43818  
A:Molecule type: mRNA  
A:Residues: 1-714 <AVI>  
A:Cross-references: GB:X59698  
R:Eisinger, D.P.; Serrero, G.  
submitted to the EMBL Data Library, June 1992  
A:Reference number: S24942  
A:Accession: S24942

A:Molecule type: mRNA  
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>  
A:Cross-references: EMBL:212608  
R:Heiserfmann, G.J.; G111, G.N.  
J. Biol. Chem. 263, 1152-1158, 1988  
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated  
A:Reference number: A28941; MUID:88330814; PMID:3138233  
A:Accession: A28941  
A:Molecule type: protein  
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,  
R:Hubbs, M.L.; Dunn, A.R.; Alexander, W.S.  
Submitted to the EMBL Data Library, April 1994  
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor  
A:Reference number: S45325  
A:Accession: S45325  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-971, 'K', 973-1210 <VER>  
A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CA55587.1; PID:9488831  
R:Paria, B.C.; Day, S.K.; Andrews, G.K.; Day, S.K.  
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993  
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse  
A:Reference number: 149643; MUID:93126380; PMID:7678348  
A:Accession: 149643  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 12-20, 22-132 <RES>  
A:Cross-references: GB:L06864; NID:9193001; PIDN:AA553029.1; PID:9567201  
C:Genetic8:  
A:Gene: EGFR  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop  
F.1-24/Domain: signal sequence #status predicted <STG>  
F.648-670/Domain: transmembrane #status predicted <TMM>  
F.712-977/Domain: protein kinase ATP-binding motif  
F.720-728/Region: protein kinase ATP-binding motif  
F.680, 695/Binding site: phosphate (Thr) (covalent) #status experimental  
F.697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental  
F.993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental  
F.1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental  
F.1157/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.0%; Score 3052; DB 2; Length 1210;  
Best Local Similarity 49.0%; Pred. No. 2, 7e-123;  
Matches 623; Conservative 169; Mismatches 370; Indels 110; Gaps 23;

11 LLLALLPQAA--STQVCTGDMKLRPASPTHLDMRLHYOGQVVOGNETLPTN 68  
14 LLLALCAAGALEEKVKVCGTSTNRLTOLGTFEDHFLSLQPMYNCEVVLGNLEITYVQRN 73  
69 ASLSFLDIOEVQGVVLAHNOVROVPLORLRIVRGTOLEFEDNYVALVINDDPLNNTTP 128  
74 YDLSLKTIOEVAGVYVLAHNTVERIPLENLOIRGNALYENTYVALAISN----- 124  
129 VTGASPGGLRELQSLTEILKGVLIQRNPOLCYODTILMKDI----FKKNQALTLI 184  
125 -YGTNRGTGLRELPMNLDELILGAVRFSNNPILCMMDTIQWRDIQVNFMSMSMDL--- 180  
185 DTNSRACHPGSPMKGSRGCESESDQSLTRTYCAGCA-RCKGPLYQYIKANSKRTIG 243  
181 -QSHPSRCPKCDPSPGNSCMGCGGECNCKLTKIICAOQCPSHRCGRSPSDCHNQCAAG 239  
244 ITELGHSDCLACLNHNSGICELHCPALVTYNTDIFESMPNREGRYTGCATCPYNY 303  
240 CTGPRESCLVCKQKQDEATCTKOTCPPLMLNPTTYQQDDVNNEGKYSFATCTVKKCPRY 299  
304 LSTDVGSCTLVCPLNNOETADGTQRCCKSPCARVYCYGMEHLREVRATSIQOE 363  
300 VVTDHGSCTVACGPRYEV-EDGIRKCKKCGPRKVCNGIGIEFDTLSINNTIKH 358  
364 FAGCKKTIGSLAFLPESPDGPASNTAPLOPQOLQVFTLEITTYGLYISAMPDLPLIS 423  
359 FKYCTAISGDHLTLVAFKGSFTRTPLDRELEILKTKVETIGTFLDIQAMPDMWTDLH 418

424 VPQNLQVTRGLTHNGAVSLTLOGISIMLGRSLRELGLALIHNNHLCGFVHTVPMD 483  
419 APENLLEIRGRKQKQGFSLAVGLNITSLGRSLKEISDGVITISGNNLTYANTINAK 478  
484 QLFRRPHOALLHTANRPEDECVEGLACIQCARGHCWCPGTQCNCSQFLRGCEVKE 543  
479 KLFGTPNQKTKIMNRAEKDCAVNVHVCNPLCSSGCMQPERDVCSCNVSRGNECVER 538  
544 CRVLQGLPREYVNAHCLPCHPECPONGSVTCFPEADQVACAHYKDPFCVACRPGC 603  
539 CNILEGEPRFEVNESECIOCHPECLPQAMNITCTGCRGPNICQAHYIDGPHCVTKCPAG 598  
604 VKPDLSTVPIKPFPEBEACQCPPTNCHSCVYDLDKQGRAGRASPLTSYSAVVGILL 663  
599 IMGENNTL-VMKYADANVCHLCNCTYCGAGPQLQCEVWPSGPKIPSLITGVIGLL 657  
664 VVVLQGVFGL-LIKRQOKIRKYTRRLLOTELVEPLTPSGAMPNOAMRIKETELRK 722  
658 FIVV-VALGIGLPMRRRIHVKTIRLLQRELEVEPLTPSGEARNQALRLIKETERFK 716  
723 VKVLGSAFGTYKGIWIPDGENVKI PVAIKVLRNTPSKANKELIDEAYVAVAGSPVY 782  
717 IKVLGSAFGTYKGLWIPDGENVKI PVAIKELREATSPKANKELIDEAYVAVAVONPVY 776  
783 SRLGICLTSTVQVLTQMLPVCCLLDHRENRGRGLGSDLLNWCQOIAKMSYLEDRVLY 842  
777 CRLGICLTSTVQVLTQMLPVCCLLDYREHNDNGSOVLNWCVOIAKMSYLEDRRLV 836  
843 HRDLAARNVLYKSPHNVKITDPLGLARLIDIDETEHADGKVPIMMALESILRRRFTHQ 902  
837 HRDLAARNVLYKTPPHVKITDPLGLAKLGAERKEHAEGKVPIMMALESILHRYTTHQ 896  
903 SDVMSYGVTVLWELMTFGAKPYDGPAREIPDLLEKGERLPPOPCTIDVYIMVKKWMD 962  
897 SDVMSYGVTVLWELMTFGSKPYDGPASDISILBGERLPPOPCTIDVYIMVKKWMD 956  
963 SECRPRFELVSESRMARDPQRFVYIQ-NELQASPLDSFTFYSLLEDMDGLVME 1021  
957 ADSRKPFELVSESRMARDPQRFVYIQDESMHLPSPDLSFYALNDEEMEVVDAD 1016  
1022 EYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLPESEEARSPPLARSEGA 1081  
1017 EYLVPOQGF-----NSBST-----SRPLFLSSLSA 1042  
1082 GSDVFDGLGKCAKAGLQSLPTHDPSPLQRYSEDPVLPSET--DGVAAPLTCSPOPEY 1139  
1043 TSN-----NSTVACINRNGSCRYKEDAPLQRYSSDPTGAVTEDNIDDAFL-----PVPEY 1092  
1140 VNQPPVRRQPPSPRSGPLPAARPAATLEKATLSFGKNGVYKDVAFAGAVENPEYL-T 1198  
1093 VNQ-SVPRKPAQSVONPVYHNPPLHP-----APGRDLAYON--PHSNAVGNPEYLMT 1141  
1199 PQGGAAPRPPPPAPFADNLVYWDQ-----DP-----PERGAPSTFGRT 1242  
1142 AQ-----PTCSSGFSNPAIMIQGSHQMSLDNDPYODFFPKETKPGITRG-PT 1191  
1243 AENPEYGLDVP 1254  
1192 AENPEYLRVAP 1203

RESULT 6  
TVCHLV  
epidermal growth factor receptor precursor - chicken  
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
C:Species: Gallus gallus (chicken)  
C:Date: 28-Feb-1996 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000  
C:Accession: A27720; A00643  
R:Max, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Venn,  
Mol. Cell. Biol. 8, 1970-1978, 1988  
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in  
A:Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720  
 A:Molecule type: mRNA  
 A:Residues: 1-1223 <LAX>  
 A:Cross-references: GB:M20386  
 R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M  
 Cell 41, 719-726, 1985  
 A:Title: c-erbB activation in ALV-induced erythroidblastosis: novel RNA processing and pro  
 A:Reference number: A00643; MUID:8522822; PMID:2988784  
 A:Accession: A00643  
 A:Molecule type: mRNA  
 A:Residues: 585-1223 <NIL>  
 A:Cross-references: GB:M10066  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor  
 specific protein kinase  
 F:1-30/Domain: signal sequence #status predicted <SIG>  
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>  
 F:31-654/Domain: extracellular #status predicted <EXT>  
 F:81-307/Domain: EGF receptor extracellular domain repeat <EB1>  
 F:397-610/Domain: EGF receptor extracellular domain repeat <EB2>  
 F:655-677/Domain: transmembrane #status predicted <TM>  
 F:678-1223/Domain: intracellular #status predicted <INT>  
 F:729-984/Domain: protein kinase homology <KIN>  
 F:727-735/Region: protein kinase ATP-binding motif  
 F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #  
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:754/Active site: Lys #status predicted  
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.7%; Score 3031.5; DB 1; Length 1223;  
 Best Local Similarity 48.0%; Pred. No. 2e-122;  
 Matches 622; Conservative 174; Mismatches 356; Indels 145; Gaps 25;

8 RMGLLLALPPGAA-----STOVCTGDMKRLPASEPETHLDMRLHYOGCOVQNGLE 61  
 13 RGAALVLLVLLGVALLCSAIVEKVCQGTNNKLTQIGHVEDHTSLQRYMNCVLLSYLE 72  
 62 LTVLPTNASTLFLODIOEVGVVLIANQVRQVLPORLRIVRGTQLEFDNYALAVLDNGD 121  
 73 ITVEHNRDLTFELKTIOGVAGVLIAMNDVIFLENQIIRGNLVYNSRLAVALSYH 132  
 122 PLNNTTPVSGAPGLRELRLSLTEILKGGVLIQRNPOLCYODITLWKDIFHKMQLAL 181  
 133 -MNTQ-----GLRELPMKRLSEILNGVKISNNPKLCNMDVLNNDIIDSRL-PL 182  
 182 TLID-TNRSRACHPCSPWCKSRCKGSESDCOGLTRIVCAGGCA-RCKGPLQYITKANS 239  
 183 TVLDFASNLSSCPKCHPCTEDHCWAGEONCOTLTQVICAQCSGRCRGVVPSCCHNQ 242  
 240 KFIGITELKHSDDCLAFHNSGICELHPALVYNTDTFESMPREGRYFGASCVTAC 299  
 243 CAAGCTGPRESDCLACRFKRDATCKOTCPPLVLYNPTTYQMDVNPBESKIFGATVREC 302  
 300 PYNLTSTDVSGCTLVCPRLNQEVTADGTQRCCKSKPCARVYCYGLGMEHLREVAVYSA 359  
 303 PHNVTVTHDGSVCVASCNDTYEV-ENGVKCKKCDGLCSKVCNGIGIGELKGLISNAT 361  
 360 NIOGFACCKTFGSLAPLPESFDDDPASNTAPLPBOLQVETLEITGYLYISAMPDSL 419  
 362 NISFKNCTKINGVSTILPVAFLDAPFKTLPBPCKLDVRYVEISGFLIIAMPDNA 421  
 420 PDLISFONLOVIRRIIANGAYSLTLOGISMTGLSLRELSGLLIIHNTHLCFVHT 479  
 422 TDLYAFENLEIRKTRQHOGYSLAVVNLKIOSGLSLKSIISGDIAIMOKKYLCTADT 481  
 480 VPMDOLEFRNPHQALLHTANREDECVBGLAQHOLCARGHCWGPGPTQVNCOSQFLRQE 539  
 482 MNMRSLFATOSQKTKIIIONRNKNDCTADRHVCDEPLCSVGCWCGPFCFCFRFSRQKE 541  
 540 CVEECRLVGLPRRYVNAHGLPCHPEQPNG---SVTCGPRAQDCVACAHYKDPFC 596

Db 542 CVKQCNIIQGEPRERERDSKCLPCHSECLVONSTAYNTTCSGPGPDHCKCAHFLDGPHC 601  
 597 VARGSGVAPDLSTVYPIKFPDEGACOPCPINCTHSCVDUDDKCGPAGRASPLTSYS 656  
 602 VKACGAGVIGENDTL-VKRYDAANAVCOLCHPNCRTCKGPGLEGP--NSGKTPSIA 657  
 657 AVV-GILLVVVLGVVFGLIKRQOKIRRYTRRLLOETELVEPLTPSGAMPQOAMRL 715  
 658 GVGGGLCLLVVVGIGIGLYLRR-HVKKRTLRLQRELVEPLTPSEANQNHRL 716  
 716 KETELARKVVLGSAFGVYVYGIKIPDEBNVKIPAIKYLRENTSPKANKELDEAYVMA 775  
 717 KETEFKVKVLSGSAFGVYVYGLMIPBEKVKIPAIKELREATSPKANKELDEAYVMA 776  
 776 GVGSFYVRLGICLTSTVOVLTOAMPGLLDHRENGRGLSGODLLMWCQIAKMGY 835  
 777 SVDNHVRRLGICLTSTVOVLTOAMPGLCLDYREHNDNGSYLLMWCQIAKMGY 836  
 836 LEDVRLVHRDLAARNVLKVSBNHVKITDPGLARLIDIDETEHADGKVPIMKMALESL 895  
 837 LEERLVHRDLAARNVLKVTQOHVKITDPGLAKLIGADEKEVHAEGKVPIMKMALESL 896  
 896 RRRFTGSDVWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955  
 897 HRIYTHOSDWSYGVTTWELMTFGSKPYDGIIPASEISSVLEKGERLPQPICTIDVYIM 956  
 956 VKCMNIDSECPREPELVEFSRMAARDPQRFVVO-NEDLGPASPLDSTFYSLLEDMM 1014  
 957 VKCMNIDSECPREPELVEFSRMAARDPQRFVVO-NEDLGPASPLDSTFYSLLEDMM 1016  
 1015 GDLYAEVYLPOOGFFPCDPAPGACGAWHHRHSSSTRSGGDTLLGLEPSEEAAPRP 1074  
 1017 EDYDADVELVPHQGF-----NSPST-----SRIP 1042  
 1075 L-----APBEGASDVPFDGLAMGAKGLSLPTHDPSPLORYSEDPVLPSET--DGY 1127  
 1043 LLSLSATSNNASATCID-----RNGQGHVPEDESFVQRYSSDPGNFLEESIDGQF 1094  
 1128 VAPLTGSPQRYVNOVDVRPQPSPREBRLPAARAGATLBRATKLSGCKGVAVDV-- 1185  
 1095 L-----PAPRYVNO--LMPKPS-----TAAVQNOIYNNISLT 1125  
 1186 -----AFGAVNPEVLTPOGGAAPQHPAPFAPFADNYVMQ----- 1225  
 1126 AISKLPMSRYONSHSTAVNDPEYL-----NTNOSPFLAKTYPESSPYLIQSGHNOIN 1177  
 1226 -DPE-----RQAPSTFGKTPAENPEYLGIDVP 1254  
 1178 LDNPDYOODPLPNETKPMGLLKVPAENPEYLRVAAP 1214

Db 1226 -DPE-----RQAPSTFGKTPAENPEYLGIDVP 1254  
 1178 LDNPDYOODPLPNETKPMGLLKVPAENPEYLRVAAP 1214

RESULT 7  
 A47253  
 epidermal growth factor receptor, HER4 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
 A:Accession: A47253  
 R:Piowman, G.D.; Culousoou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;  
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
 A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderm  
 A:Reference number: A47253; MUID:93189574; PMID:8383326  
 A:Accession: A47253  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1308 <ELO>  
 A:Cross-references: GB:I07868; NID:9337359; PIDN:AA59446.1; PID:9337360  
 A>Note: sequence extracted from NCBI backbone (NCBI:126842)  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor  
 F:716-981/Domain: protein kinase homology <KIN>  
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 42.9%; Score 2909.5; DB 2; Length 1308;  
 Best Local Similarity 44.9%; Pred. No. 3.6e-117;  
 Matches 605; Conservative 182; Mismatches 386; Indels 173; Gaps 28;

```

Oy 9 WGLLIALLPPGAA-----STOVCTGTDMLRLPASPEHLDMLRHLYOGCOVVOGNIELTY 64
Db 8 WWWVSLVLAAGTVPQSDSQSCAGTENKLSLSDEQQYRALRKYYENCEVVMGNILETS 67
Oy 65 LPTNASLSFLDIOIOGVGVLLIAHNOVROVLORLRIYNGTOLFEDNYALALVDNDPLN 124
Db 68 IEHNDSLFSLRVREVTGYVALNQFRLPLENRIIRGTLYLVEDRYALALFLYVRKDQ 127
Oy 125 NTTPTVGTASPGGLRELOLRSLTEILKGVLIQNPOLCYODTILMKDIEFHKNQALATLI 184
Db 128 NF-----GLQELGLKNLLEILNGVYVQDNKFLCADTIIHMODIYRANWPNLTLIV 178
Oy 185 DTNRSRACHPCSPWCKGRSCWGESSEDCSLTRTYCAGGC-ARCKGPLFOYIKANSKFTG 243
Db 179 STNGSSGGRCHKSCGTG-RCWGPTEHNCOTLTRIVCAEQDCRCYGPVSDCHRECAAG 237
Oy 244 ITELKHSQCLACLFHNHSGICELHCPALVTYNTDFESMNPREGRTFGASCVTACPNY 303
Db 238 CSGPDTDCPACMNFNDSACVTQCQTFVTYVPTTQLEHNFNAKTYGAFCKKCPHNF 297
Oy 304 LSTDVGSCTLVCPILHNOEVTAEEDGTORCEKSCPKARVCYGLGMEHLREVRATVSANIQE 363
Db 298 V-VBSSSCVRACRSPSKMEV-EENGIMKCKPCTDIPCXADGIGTGSLSAQVDSNDIK 355
Oy 364 FAGCKKIFGSLAFLEPSFDGPASNTAPLOPQLOVFEETLITGYLYISAMPDLPUS 423
Db 356 FINTCKINGNLIFLTGTGHDGPDYNAIEADPEKLVNFRVREITGTLNQSMPNMTDS 415
Oy 424 VFQNLQVIRGRILHNGAVSLTQGLISMLGLRSREISGLALHNNHLFCVHTVMPD 483
Db 416 VFSNLVTITGGRVLYSGLSLILKQOQITSLQFOSLEISAGINITYDNNLCYHTINNT 475
Oy 484 QLFRRNPQALLHTANRPEDECVGEGLACHQLCARGHCMGPRPTQCVNCSQFLRGCEVBE 543
Db 476 TLFSTINQRIIVIRDRKAKENCTAEGVNCNHLCSDDCGMGPGRDQCLSCGRFGRGICIES 535
Oy 544 CRVLOGLRREYVNAHCLPCHPECOP-QNGSVTCGPEPADQVACAHYDPPFCVACRCS 602
Db 536 CNLYGGEFREPEFENGSIQVECDPQCEMDEGILLTCHGPGDNCCTKSHFDXGPNCEKCPD 595
Oy 603 GVKPPLSYMPWKFPDEBACOPCPINTGSCVDDDKG-----PAEVRASPL 651
Db 596 GLQGANSF--IFKYADPRECHPCPNCTQCGNGPNSHDCTIYPTWGHSTLPQHAH-TPL 652
Oy 652 TSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKXTMRRLQETELVEPLTFSGAMPNQA 710
Db 653 --IAAGVIGGLFILVIVGLTFAVVYRRKSIK-KKQALRRFL-ETELVEPLTFSGAPNQA 708
Oy 711 QMRILKETELRKVKVLGSGAFGTYYKGIPIGGENVKIYVALKVLRENTSPKANKEIIDE 770
Db 709 QRLIKETELRKVKVLGSGAFGTYYKGIWVPEGETYKIPVALKILNETTGPANVEFMDE 768
Oy 771 AYYMAGVSPYVSRILGICLTSTVOLVNTQMPYGLLDHVRNBRGLSGODLLNMCQIA 830
Db 769 ALIAMSMDHPLVRLGLVCLSTPILQVLTQMLPHGCLLETVHEHKDNISQSLLLNMCVOIA 828
Oy 831 KGMSTLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKMA 890
Db 829 KGMSTLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKMA 888
Oy 891 LESILARRFTHOSDVWVGYVWELMTGAKPYDGIIPAEITDILLKGRRLVQPPICITD 950
Db 889 LECHYRKFTHOSDVWVGYVWELMTGAKPYDGIIPREIPDLLKGRRLVQPPICITD 948
Oy 951 VYIMVKKCMIDSECRPFRELVSSEFARMADPQRFVVIQNEED-LGPASPLDSTFYRSLL 1009
Db 949 VYIMVKKCMIDADSRPFKELAAEFSSRARKDPQRYLVYIQGDRKMLSPSPNDSKRFQNL 1008
Oy 1010 EDDMDGLVDAAEYVLPQOGFFCPDPAPAGAGMVHHRHSSTRSGGGLDILGLEPSEBE 1069

```

```

Db 1009 DEEDLEDDMDAAEYVLP-QAFNIPPP-----IYNSRARIDSNRS-----ELGHSPPRAY 1056
Oy 1070 APRS-----PLAP-SRGAGSDVFPDGLGCAKGLQS 1100
Db 1057 TEMSGNQFYRRDGGFAEGVSVPYRAPRTSTIPEAPVNOGAFAEIFDSCCGCTLRKPPA 1116
Oy 1101 LPTHDSPLORYSEDPVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRPQPSPR 1153
Db 1117 PHVQEDSSTQRYSDPTFAPAPERSPRGELIDEGTTPMKDKKQETLYNVE----- 1167
Oy 1154 EGPLPAPAPAGATLERATLSPGKNGVNDVAFGAVENPEYLTPOGGAQAPQHPHPPA- 1212
Db 1168 ENPFVFSRR-----KNGDLD-----ALDNPETHNASNC-----PPKAE 1199
Oy 1213 -----FSPAPFNLYYWDODPPERGA--PPSTF 1237
Db 1200 DEYVNEPLYNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDMVHNSLPPRSTLQHPDYL 1259
Oy 1238 KGTPT-----AENPEYL 1249
Db 1260 QEYSTKYRYKQNGRIPIVAENPEYL 1285

```

## RESULT 8

S06142  
 protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish  
 Maltetracte names: epidermal growth factor receptor homolog; kinase-related transform:  
 C1Species: Xiphophorus maculatus (southern platyfish)  
 C1Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000  
 C1Accession: S06142; S13809  
 R.Mitbrodt, J. Adam, D. Malitschek, B. Maueleler, W. Raulf, F. Telling, A. Robe:  
 Nature 341, 415-421, 1989  
 A1Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu  
 A1Reference number: S06142; MUID:90015140; PMID:2797166  
 A1Accession: S06142  
 A1Molecule type: DNA  
 A1Residues: 1-1166 <MIT>  
 A1Cross-References: EMBL:X16891; NID:965290; PIDN:CAA34770.1; PID:965291  
 R1Adam, D.; Maueleler, W.; Schartl, M.  
 O1cogene 6, 73-80, 1991  
 A1Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph  
 A1Reference number: S13807; MUID:91125882; PMID:1846957  
 A1Accession: S13809  
 A1Status: preliminary; translation not shown  
 A1Molecule type: DNA  
 A1Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>  
 A1Cross-References: EMBL:X56319; NID:965284; PIDN:CAA39763.1; PID:965285  
 C1Genetics:  
 A1Gene: mrk  
 A1Map position: Y  
 A1Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1  
 C1Superfamily: epidermal growth factor receptor; protein kinase homology  
 C1Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t  
 F11-25/Domains: signal sequence #status predicted <SIO>  
 F126-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>  
 F1707-972/Domains: protein kinase homology <KIN>  
 F1715-723/Region: protein kinase ATP-binding motif

Query Match 38.4%; Score 2609; DB 1; Length 1166;  
 Best Local Similarity 44.8%; Pred. No. 2.2e-104;  
 Matches 567; Conservative 162; Mismatches 400; Indels 138; Gaps 28;

```

Oy 4 AALCRWGLIALLPPGAAS-----OVCTGTDMLRLPASPEHLDMLRHLYOGCOVVOGNI 59
Db 8 AALQD--LLLVLSISRCSTDPDRKVCQGTSMQMTM---LDNHVYKMKMKMYSGCUNVLEN 62
Oy 60 LELTYLPTNASLSFLDIOIOGVGVLLIAHNOVROVLORLRIYNGTOLFEDNYALALVDN 119
Db 63 LEITTYQENODISFLQIOGVGVLLIAHNEVSTIPLVNLIRIKQNLVYEGNFTLLVWSN 122
Oy 120 GDPLNNTPTVGTASPGGLRELOLRSLTEILKGVLIQNPOLCYODTILMKDIFHKNQNL 179

```



```

Db      123 YOK-NPSP--DVTQVGLKQULSNTLEILSGVKVSHNPLLCNVETINMMDIVDKTSP 179
      180 ALTLIDNRSRACHPCSPMCKSGRCWSESDCSLTRTVAGGC-ARCKGRLPOLYIYAN 238
      180 TMLNIPPAFERQCKCHGICVNSGWAIPGHCKQKFTLLCAEQCNRCRCKPRLDCNE 239
      239 SKFGLITELKSDCLCLHFNHSGICELHCPALVTYNTDTESMPNEGRYTFGASCYTA 298
      240 HCAGCGTGPRAITDCLACGDFNDGCTCKTCCPPKTYDIVSHQVVDNPNIKYTFGAACYKE 299
      299 CPVYLYSTDVGSCTLVCPLEHNEVTAEDGTORCEKSGPCARVCGIGMEHLREVRATYS 358
      300 CPSPYVYTE-GACVRSACMLLEV-DENKSGCKPCDGVCPKCDGIGISLSTIANVS 357
      359 ANIOEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOEOLQVFETLEITGYLYISAMPDS 418
      358 TNIRSFNCTKINGDITLNRNSFEGDPRYKIGTMDPEHLMNLTYYKEITGYLVIMMHPEN 417
      419 LPDLVSQNLQVIRGRLHNGAYS-LTLQGLIGISMLGRSIRELSGSLALIHNTHLCEV 477
      418 MTSLSVFQNLLEIRGRITFSRGSFVVQVHQLMGLRSIKESVAGVILIKNTLQLRYA 477
      478 HTVPMDOLEFRPHOALLHTANRPEDECVEGLACHQICARHCWGPRGTQCVNCSQPLRG 537
      478 NTIMWRLEFRSESDSIEYDART-----ENQTCNNECSDEGCM-PGFTVCVSCILHYDRG 529
      538 QECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCAACHYKOPFECV 597
      530 GRCVASCNLLQGEEREAQVORCVQCHOECLVQDLSITCYPRGANCSSKSHFDGQCI 589
      598 ARCPGSKPDLSTYPIKFPDEBACOPRLNCHSCVLDLKKCPABQRPASPLTYSIVA 657
      590 PRCPHGILGQDGLT-INKYADIKMGQCPCHQNTQSGSPGLSGCRDP-IYSHSLAVGL 647
      658 VVGLIVVVGAVGVLIKRQOKIRKYTWRLLOETELVPLTPSGAMPQAOARILKE 717
      648 VSGLLIVIVALLVLLRRRRRIK-RKRTIKCLQEBELVEPLPSQAPQOALRIKE 706
      718 TELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKAKEILDEAYVAGV 777
      707 TEFKKQDVLSGAFGVYKGLMNDGENIRIPVAIKVLRNTPSKAOEVLDEAYVAVS 766
      778 GSYVSRLLGICLTSTYQVLTQMLPYGCLLDHRENRGLSGODLWMCQIAKMSYLE 837
      767 DHPHVCRLLGICLTSAVOLVTOMLPYGCLLDYVQHOERICGQWLNMVQIAKMSYLE 826
      838 DVLVHRDLAARVNLVSPVHVKITDGLARLIDIDETEHADGCKVPIWMALLESILR 897
      827 ERHLVHRDLAARVNLVKNPVHVKITDGLSKLLTADKEYQADGCKVPIWMALLESILR 886
      898 RFTGSDVMSYGVTVWELMTFGAKPYDGIPIAREIPIDLLEKGERLPOPICTIDVYIMTWK 957
      887 TYHGSDVMSYGVTVWELMTFGSKPYDGIPIAKEIASLWENGERLPOPICTIEYIMILK 946
      958 CMWIDECRRFRELIVSEFSKMAADPORFVYIIONEDIGPASPLDSTFYRSLLEDDMGDL 1017
      947 CMWIDSSRRFRELIVGEFSQMARDPSSRYLVIQ---NLPSLSDRLFSRLSSD--DV 1001
      1018 VDAEETLVPOQGFPCDPAPGAGGVMHHRSSSTSGGDLTLGLEPSEEARPSRLAP 1077
      1002 VDADELTLPYKRI-----NROGS-----EPCIP 1024
      1078 SEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEDPV-DLPSETDGYVAPLTCSPQ 1136
      1025 PTGH-----PVRENSITLRNLSIDPTQNLLEMDLQD----- 1055
      1137 PEYVNPQVRRPQ-----PSPRE-----GLP-ARAPAGATLEAKTILSPKNGVYKD 1183
      1056 -EYVNPQGETSRRLSDIYNPNEDLTDGWPVLSLSOEALTNFSPEYINTQNSL--- 1111
      1184 VFAFGAVENPEYLTPOGGAAPQHPRPAPSPAFDLVYWDQDPRBGRAPPSFTFKGTPPA 1243
      1112 PLVSSGSMDDPY---QAG-----YQAAV-----LPQYTGALTNGMFLPAA 1149

```

```

      1244 ENPEYIG 1250
      1150 ENLEYIG 1156

RESULT 9
A36223
kinase-related transforming protein (erbB) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; 159164
R:Kinsman, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <RFA>
R:Kinsman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-
A:Reference number: 159164; MUID:90311312; PMID:2164210
A:Accession: 159164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559,'G',561-957,'F',959-1063,'G',1065-1342 <RES>
A:Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match      34.5%; Score 2345.5; DB 2; Length 1342;
Best Local Similarity 40.0%; Pred. No. 4.6e-93;
Matches 525; Conservative 190; Mismatches 467; Indels 129; Gaps 32;

      10 GLLALLPFGAA--STQVCTGDMKRLRPSPEHLDMLHLYOGCGVQGNLELYLPT 67
      11 GLLPFLARSGSEVNGSQAVCEPTLNGLSVTGDAENQVOTLYKLYERCEVWNGNLEIYLTGH 70
      68 NASLSFLDIOIEVQGYVLIANQVROVPLQRLIRVIRGTOLFEDNYALAVLDNGPLNNTT 127
      71 NADLSFLQWIREVTGYLVAMNEFSTLPRLNLRVYRGTQVYDGKFAIFVM---LNYNT 125
      128 PVTGASPGGLRELQLSLEILKQGVLIQRPOLCYODTILMKDIFHKNNQLALLTIDTN 187
      126 ---NSSHALRQRLRLQVLTIELISGVYIEKNDKLCHMDITDMDIYVARDR---AEIYKD 178
      188 RSRACHPCSPMCKSGRCWSESDCSLTRTVAGGC-ARCKGRLPOLYIYANSKFGLITE 246
      179 NGRSCPCHCEVCKG-KCWGGSSEDCQTLTTICAPQCNCHCFBPNPQCHDEACGCGSG 237
      247 LKHSDCIACILHFNHSGICELHCPALVTYNTDTESMPNEGRYTFGASCYTACRYNYLST 306
      238 PQTDFCFACRHFNDSACVPRCPQPLVYNKLTQLEBNPTTKYQYGGVCAVSCPHNFV-V 296
      307 DVGSCTLVCPLEHNEVTAEDGTORCEKSGPCARVCGIGMEHLREVRATYSANIOEFAG 366
      297 DQTSVYRACPRPDMMEVD-KNGLKMCBPCGGLCPKACBGGSG--SRFQYVDSNNINGFVN 353
      367 CKKIFGSLAFLEPSFDGDPASNTAPLOEOLQVFETLEITGYLYISAMPDSLPLDSVFQ 426
      354 CTKILGNLDFLLTGLNGDPMHKIPALDPEKLVNFRVREITGYLNIQSWPRPHNHNSVFS 413
      427 NLQVIRGRILHNGAYS-LTLQGLIGISMLGRSIRELSGSLALIHNTHLCEVYTVWDOU 485

```



Db 414 NLTTIGRSIYNRGFSLLIMKLNVTSLGFRSLKEISAGRIYISANROLCVHHSLMNTKV 473  
486 FRNPHOALLHTA-NRPEDCVGEGGLACHOLCARGHCMGPGPTQCVCNCSOPLRGQECVEEC 544  
474 LRGTLEERLDIKHNRPRDCVAEGKVCPLCSSGCMWPGQCLSCNRYSRGVCTVHC 533  
545 RVLOGLPREVYNAHRLCLPCHPECOPBONSVTCTCFGBEADCVACAHYKDPFCVAPCRSGV 604  
534 NFLNGEPREFAHAEACFCSCHEPCQMEGTATCNGSGSDTCAQCAHFRGPHCVSSCPHGV 593  
605 KPDLSYMPIMKFPDEEGACQCPINCTHSCVDLDKCGPAEGR-----SPLTSIYSAVVG 660  
594 LG--AKGPIYKYPDVQNECRCHENCTCGCKGPELDGLGVLIGKTHLTMLATVAG 651  
661 ILVVLGVVFGILIKRQOKIR-KYTRRLLOETELVEPLTPSGAMPNOQMRLKETE 719  
652 --LVVIFMMLGSTFLYMGGRIONRARMRYLERGESIEPLDPS-EKANKYLARIFKETE 708  
720 LRKVKLGSGAFGTYVKGIMI PDGENVKI PVAIKVLRNTSPKANKELDEAYYVAGVS 779  
709 LRKLKVLGSGVFGTVHKGWMIPEGESIKIPVCIKVIEBKSGRQSFQAVTDHMLAIGSLDH 768  
780 PYVRRLIGLCTSTVQVLTQMLPVCCLLDHYRENRGRIGSDLLMCMQIAKMSYLEDV 839  
769 AHIVRLGLCGSSSLQVLTQYLPGLSLDHYRQHRGALGPOLLNMGVQIAKMYLYEEH 828  
840 RLTVRDLAARVNLVKS PNHVKITDEGLARLLDIDETEYHADGKVPKIMMALLESTLRARF 899  
829 GMVRNLAAARVNLKSPSOVQVADGVADLLPRDDKOLLYSEAKPIKMMALLESIHFKKY 888  
900 THQSDVMSYGVTVLWELMTFGAKPYDGI PAREIPDLLEKGERLPOPPICITIDVYIMVXCM 959  
889 THQSDVMSYGVTVLWELMTFGAEPYAGRLAEVPLDEKGERLAQPOICTIDVYIMVXCM 948  
960 MIDSECRREFELVSEFSRMARDPQRFVYIQNEDLGRA---SPLDSTYRSLLEDDMGD 1016  
949 MIDNIRTEKELANEFTRMARDPRLYLVIKRES-GPGIAPGPEPHGLTNKKLEVELEP 1007  
1017 LVDAEYLVPOQGFECPPDPAAGAGMVAHRRSSSTRSGGDLTLGLEP-SEEARPSPL 1075  
1008 ELDDLDDLEAED-----NLATTLTGLSALSPVCTINLRPGSQSL 1048  
1076 APSPGAGSDVDGDLGMAKAGLOSPLTHD-PSPLQRYSEDPVPLP-----SETDGV 1128  
1049 SPSSGY-MPMNQNLGESCQESAVSGSSERCPRPVSLH-----PMRGLCLASESSEGHV 1101  
1129 A-----PLTSPQPE-----YVNOVDVAPQPPSPREG-----L 1157  
1102 TGSEAELOEKVSKCRSRSRSPRPRGDSAYHSQHSLLTPVTLSPGLEEDVNGYVM 1161  
1158 PAARPAATLERAKTLSP-GKNGVY-----KDVAFAGAVENPEYLTPOGGAAPQHP 1210  
1162 PDTHLKGTSPSREGTSLVSGSLVGLTGEDEED-----EVEYMMRRRRHSP-PHP 1212  
QY 1211 PARSAPADNLTYMD-----QDPERGAPPTSTFKGTPTAENEYLT 1249  
Db 1213 RPSLSLELGYEYMDVSGDLSASLSTGSCPLHPVIMPTAGTTPDEDEYEM 1263

RESULT 10  
JC4387  
Epidermal growth factor receptor homolog precursor - rat  
N:Alternate names: ErbB3 protein; HER3 protein  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 17-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Nov-1998  
C:Accession: JC4387  
R:Helixer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.  
Gene 165, 279-284, 1995  
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.  
A:Reference number: JC4387; MUID:96096535; PMID:8522190  
A:Accession: JC4387  
A:Molecule type: mRNA  
A:Residues: 1-1339 <HEL>

A:Cross-references: GB:U29339; NID:G915389; PID:G915390  
A:Experimental source: liver  
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residu  
C:Comment: This protein is a functional heregulin receptor that transduces signals to  
C:Genetics:  
A:Gene: ErbB3  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>  
F:640-659/Domain: transmembrane #status predicted <TM>  
F:705-970/Domain: protein kinase homology <KIN>  
F:713-721/Region: protein kinase ATP-binding motif  
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)  
Query Match 33.3%; Score 2260.5; DB 2; Length 1339;  
Best Local Similarity 40.1%; Pred. No. 2e-99;  
Matches 515; Conservative 170; Mismatches 443; Indels 155; Gaps 34;  
QY 3 LALCRWGLLALPPGAA---STQVCTGDMKRLPASPETHLDMRLHYOGCOVQGN 59  
Db 7 LOVLC----FLSLARGSEKNSQAVCEPTGLANGSLVTDADNOYOTLYKLEKECEVWGN 62  
QY 60 LELTYLPTNAGSLSPLODIOEVQGYVLAHNOVROVPLQRLIRVQTOUFEDNYALAVLDN 119  
Db 63 LEIVLTGHMADSLFQWIREVTAYVLAHNEFSVLPLENLAVVGTQYVDGKFAIFVM-- 120  
QY 120 GDPANNTPTVGAAPGGLRELQSLSTELKGVLIQNPOLCTODITLMKDIFPKNNOL 179  
Db 121 ---LNYNT---NSSHALRQLKFTOLTLEIGGVYIEKNDKLCMDITIDMDRDIYVR-- 170  
QY 180 ALTIDITRSNACHPSCPMCKSGRCWGESSEDCSLTRTVACAGC-AACKGKPOLYIYAN 238  
Db 171 GAETVKNKANGACPPCEHCVKG-RCMGPGDDCOLITITCAPQCNCRFGENPQOCHD 229  
QY 239 SKFIQITELKASDCLACLFHNSGICELHCPALVTYNTDESMNPREGRTFGASCYTA 298  
Db 230 ECAGCGSGPODTCFACRFNDSDGACVPRCEPLVYNNLTQLENNPHTKYOYGVCAAS 289  
QY 239 CPYNYLTSDVSCTLVCPYLNQEVTAEDGTORCEKSAFPCARVCYGLMEHLREVRAVTS 358  
Db 290 CPNHFV-VDQFCVRAACPPDMREVD-KHGLMKCEPCGGLCPKAEGETGSG--SRVQYDVS 345  
QY 359 ANIQFAGCKKIFESLAFLPESFDGDPASNTAPLOPELOVFEETLEETLYGLVYSAWPS 418  
Db 346 SNIDGFVNCTKIIGNLDPLIAGLVNDPMHKIPALDPEKLVNFRVRELTGLYNTIOSWPRH 405  
QY 419 LPDLVFNQLQVIRGLIHNGAYS-LTLOGIGISMLGLRSRLRELSGLALIHNTHLCEV 477  
Db 406 MHNFSVFSNLTIGRSIYNRGFSLLIMKLNVTSLGFRSLKEISAGRIYISANQOLCYH 465  
QY 478 HTVWMDQLFRNPHOALLHTA-NRPEDCVGEGGLACHOLCARGHCMGPGPTQCVCNCSOPLR 536  
Db 466 HSLMTWRLLRGSPSEERLDIKYDRPLGECLEAGKCDPLCSSGCGMGAPGQCLSCNRYSR 525  
QY 537 GQECVCEGRVYQGLPREVYNAHRLCLPCHPECOPBONSVTCTCFGBEADCVACAHYKDPFC 596  
Db 526 EGVCVTHCNFLQSPREFVHAQCFSCAPCLPMEGSTYNGSSDCAARCARFRDGHG 585  
QY 597 VARCPGSKPDLSTYPIWKFPDEEGACQCPINCTHSC--VDLDKCGPAEQRASPLTSI 654  
Db 586 VNSCPHGIIG--AGPIYKYPDAQNECRCHENCTCGCNGEBELDDCGAGVAMLSKPHLV 643  
QY 655 VSAVVGILLVYVGVGILIKRQOKIR-KYTRRLLOETELVEPLTPSGAMPNOQMRLKETE 713  
Db 644 IAVTVG--LAVIIMLIGSFLYMGGRIONRARMRYLERGESIEPLDPS-EKANKYLAR 700  
QY 714 ILKTELRKVKVLSGAFGTYVKGIMI PDGENVKI PVAIKVLRNTSPKANKELDEAYV 773  
Db 701 IFKTELRKLVKLGSGVFGTVHKGWMIPEGESIKIPVCIKVIEBKSGRQSFQAVTDHMLA 760  
QY 774 MAGVSPYVRLGLICTSTVQVLTQMLPVCCLLDHYRENRGRIGSDLLMCMQIAKMSYLE 833

Db 761 VGSIDNHAIVRLGLCPSSSLQVLTQVYPLGSLDLHVKKHRETLRPLLNNKVOIAKGM 820

Qy 834 SYLEDVRLVHRDLAARNVLYVSPNHVKTITDFGLARLDIDETEYADGGKVPKIMMALES 893

Db 821 YLEEHSHVHRDLARNVLMKSPSQVADFCVADLPPDDQQLHSEAKPTPKIMMALES 880

Qy 894 ILRRRFTHQSDVSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953

Db 881 IHFGKYTHQSDVSVGVTVWELMTFGAEFPYAGLRILAEIPDLLEKGERLAQPOICTIDVYM 940

Qy 954 IMVCKMIDSECRPFRELVSEFSRMARDPQRFVVIQWEDLGPAPLDSTFRSLLEDD 1013

Db 941 VMVCKMIDENIRPFKEFLANEFTRMARDPPRYLVIKKAS-GPGRP--PAAEPSVLTKE 997

Qy 1014 MGDVLDAEYLVPOQGFPCPDPAFGAGVHHRHSSSTRSGGDLTTLGLEPSEE----- 1068

Db 998 L-----GEAELEPEL-----DLDDLLEEEGLATS 1023

Qy 1069 -----EARRPLA9SEG-----AGSDVPDGLGMAKAGLSLPTH 1105

Db 1024 LGSALSLPTGTLTPRGQSLSLSPSSGYPMNQSLGAECLDSAVLGGREGQSPRISLH- 1082

Qy 1106 PSPAORYSEDPVLPSETDGV-----APL-----TC-----SPOPE-----YNOQPDV 1145

Db 1083 PIPGR-----PASESSEGHVTCSEAELOEKVSVCSRKRSPRPRGDSAYHSQHS 1135

Qy 1146 RPOPPSPRECP-----LPAARPAGATLERAKTASP-GKNGV-----KDVPAF 1187

Db 1136 LRTVTPLSPRGLEEDDNGVYMPDTHLRGASSSREGLTSSVGLSVLTGTEDEDD----- 1191

Qy 1188 GGAVENBEYLTPOGGAAPQPPHP 1210

Db 1192 -----EEVEYMNKRKRGSP-PRPP 1209

## RESULT 11

TFVFLV  
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus  
N/Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase  
C/Species: avian leukosis virus, ALV  
C/Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999  
C/Accession: B00643  
R/Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.  
Cell 41, 719-726, 1995  
A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
A/Reference number: A00643; MUID:85228222; PMID:2988784  
A/Accession: B00643  
A/Molecule type: mRNA  
A/Residues: 1-698 <NTL>  
A/Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750  
A/Note: In Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal  
C/Comment: This protein is synthesized as a gag-env-erbB protein.  
C/Genetics:  
A/Gene: gag-env-erbB  
C/Superfamily: epidermal growth factor receptor; protein kinase homology  
C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific  
F/1-6/Product: gag protein (fragment) #status predicted <AG>  
F/1-59/Product: env protein (fragment) #status predicted <ENV>  
F/60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>  
F/194-459/Domin: protein kinase homology <KIN>  
F/202-210/Region: protein kinase ATP-binding motif  
F/229/Active site: Lys #status predicted

Query Match 26.0%; Score 1766.5; DB 1; Length 698;  
Best Local Similarity 52.2%; Pred. No. 1.3e-68;  
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 578 GPEADQVCAHYDPPFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPIINCHSCVDL 637

Db 60 GP--DHMKCAHFIDGPHCVACAPAGVLGENDTL-VMKYADANAVCQLCHPNCIRGCKGP 116

Qy 638 DDGCGPAREGRASPLTISVSAV-VGILLVVLGVFGLIKRQOKIKRYMRLLOETEL 696

Db 117 GLEGP---NSKTPSIAGVVGGLCLVWVGIGLGLYLR--HIVRKTLRLLQEREL 172

Qy 697 VEPLPSGAMPQAOAMRILKETELRKVVYLSGAGFYKGIWIDGENVKIPVALIKVR 756

Db 173 VEPLTPSGAPQAOALRIKLETFKVKVYLSGAGFYKGIWIDGENVKIPVALIKVR 232

Qy 757 ENTSPKANKEILDEAVYVAGVSPYVSRLLGLCLTSTVQVLTQMLPYGCLLHVENRGR 816

Db 223 EATSPKANKEILDEAVYVAGVSDNPHVCULLGLCLTSTVQVLTQMLPYGCLLHVENRGR 292

Qy 817 LGSQDLNWCQOIAKMSYLEDVRLVHRDLAARNVLYVSPNHVKTITDFGLARLDIDETE 876

Db 223 IGSQYLLNWCQOIAKMSYLEERLVRDLAARNVLTVPQVKTITDFGLAKLGADKEX 352

Qy 877 YHAQSGKPIKIMMALESTLRFRFTQSDVSVGVTVWELMTFGAKPYDGIIPAREIPDLLE 936

Db 333 YHABGKVPKIMMALESTLHRVYTHQSDVSVGVTVWELMTFGSKPYDGIIPASEISSYLE 412

Qy 937 KGERLPQPICTIDVYIMVCKMIDSECRPFRELVSEFSRMARDPQRFVVIQ-NEDLG 995

Db 413 KGERLPQPICTIDVYIMVCKMIDADSRRPFRELVSEFSRMARDPQRFVVIQ-NEDLG 472

Qy 996 PASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFPCPDPAFGAGVHHRHSSSTRSG 1055

Db 473 LPSPTDSKFYRTIMEEEDMEDIVDAEYLVPHQGF-----NSPST--- 513

Qy 1056 GGDULTGLPEBEERAPRPL-----APSEGASDVFPDGLGMAKAGLSLTPHPSPLQ 1110

Db 514 -----SRPLLSLSATSNNSATNID-----RNOGHVRSDSFQ 550

Qy 1111 RYSEDPVLPSET--DGVVAPLTCSPQPEYVNOQPDVPRPPSPREGPLPARPAGATVE 1168

Db 551 RYSDPTGNFLEESIDDFL-----PAPEYVNO--LMPKAPS----- 585

Qy 1169 RAKTLPGRKNGVADV-----AFGAVENBEYLTPOGGAAPQPPHPAP 1213

Db 586 -----TAMVQNOIYNNISLTAISKLPMDSRQYNSHSTAVDNPEYL-----NTQSPLA 633

Qy 1214 SPARDNLYYMQO-----DPEE-----RGAPSPFFKGTPTAENBEYLGDLVP 1254

Db 634 KTFEESPYWIQSGNQHQLNDNPYQDFLPNETKPNGLKVPAAENBEYLRVAAP 689

## RESULT 12

TVYUH  
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)  
C/Species: avian erythroblastosis virus  
C/Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999  
C/Accession: A00644; A38022  
R/Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.  
Cell 35, 71-76, 1983  
A/Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fan  
A/Reference number: A00644; MUID:84026539; PMID:6313229  
A/Accession: A00644  
A/Molecule type: DNA  
A/Residues: 1-604 <YAM>  
A/Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678  
R/Debutre, B.; Henry, C.; Benabissa, M.; Bisette, G.; Clavertie, J.M.; Saule, S.; Martir  
Science 224, 1456-1459, 1984  
A/Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type c  
A/Reference number: A38022; MUID:84223957; PMID:6328658  
A/Accession: A38022  
A/Molecule type: DNA  
A/Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>  
A/Cross-references: GB:K02006  
C/Genetics:  
A/Gene: erbB  
C/Superfamily: epidermal growth factor receptor; protein kinase homology  
C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific  
F/110-195/Domin: protein kinase homology <KIN>  
F/118-146/Region: protein kinase ATP-binding motif  
F/165/Active site: Lys #status predicted

Query Match 25.1%; Score 1703; DB 1; Length 604;  
 Best Local Similarity 52.2%; Pred. No. 5.8e-66;  
 Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

587 CAHYDPPPCVACRCSPGKPDLSYMPIMKFPDEEGACQPCPINCTHSCYVDLDDKCPRAQ 646  
 Db 3 CAHFIDGPHCAVCACGAVLGENDTL-VKRYADANAVCOLCHPCTCRGCGPGLCECP--- 58  
 647 RASPLTSIVSAVV-GILLVVLGVGVGILIKRQOKIRKYTRMRLLQETLEVEPLTPSGA 705  
 Db 59 NSKTPSIAGVGVGLCLLVVGLIGLYLRRR-HIVRKTRRLLOEHELVEPLTPSGE 117  
 706 MPNOQMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKI PVAIKVLRNTPSPKANK 765  
 Db 118 APNOHRLRLKETEFKVKVLSGAFGTYYKGLWIPGEKVKI PVAIKELREATSPKANK 177  
 766 EILDDAAYVAGVGSFYVSRILGICLTSTVQLVTQLMPYGCCLLDHVRNRRGLSGODLNM 825  
 Db 178 EILDDAAYVAGVGSFYVSRILGICLTSTVQLVTQLMPYGCCLLDYIREHKNIGSQYLLNM 237  
 826 CMQIAKGSYLEVDRLVHRDLAARVNLVYKSPNHVITDGLARLLDIDETEHADGKVP 885  
 Db 228 CVQIAKGMNLYLEERLVLHRDLAARVNLVYKTPQHVKITDGLAKLGADEKEVHAEGKVP 297  
 886 IKMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPP 945  
 Db 298 IKMMALESILHRITHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQPP 357  
 946 ICTIDVWIMVCMWIDSECRPRELVESEFMRMDPQRFVVIQ-NEULGASPLDSTF 1004  
 Db 358 ICTIDVWIMVCMWIDSECRPRELVESEFMRMDPQRFVVIQ-NEULGASPLDSTF 417  
 1005 YRSLEDDDMGDLVDAEYLVPQGGFCDDPAPGAGMWHHRSSSTFSGGDLTLGLE 1064  
 Db 418 YRTLMEEEDMEIYVDAEYLVPHQGF-----NSPST----- 449  
 1065 PSEEEAPRSPPL-----APSEGAGSDVFDGLGMAKGLQSLPTHPSPPLQ 1119  
 Db 450 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHPRVEDSFVQRYSSDPTCN 495  
 1120 LPSET--DGVAAPLTCSPQPEYVNOQDVRRPQPSREGPLPAARAGATLEAKTSLRCK 1177  
 Db 496 FLEESIDGGFL-----PAPEYVNO--LMPKKPSTAM----- 524  
 1178 NGVVDVPAF-----GGAIVENPEYLPQGAAPQHPAPSPAPFD 1218  
 Db 525 --VQNGIYNFISLTAISKLPMSRYQNSHSTAVDNPYL-----NTNOSP LAKTYFE 574  
 1219 NLTYWDQDPPERGAPSPSTFKGTPTAENPEY 1248  
 Db 575 SSPYWIQSGNHQ-----INLDNPDY 594

RESULT 13

S35745  
 C:Species: avian erythroblastosis virus  
 C:Date: 03-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 28-Feb-1997  
 C:Accession: S35745  
 R:Vennstrom, B.  
 submitted to the EMBL Data Library, March 1993  
 A:Reference number: S35743  
 A:Accession: S35745  
 A:Molecule type: DNA  
 A:Residues: 1-544 <VEN>  
 A:Cross-references: EMBL.X12707  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific F  
 F.135-400/Domain: protein kinase homology <KIN>  
 F.143-151/Region: protein kinase ATP-binding motif  
 F.170/Active site: Lys #status predicted

Query Match 24.3%; Score 1647; DB 2; Length 544;  
 Best Local Similarity 54.9%; Pred. No. 1.3e-63;  
 Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

578 GREADQCAAHYKDPCEVCARCPGKPDLSYMPIMKFPDEEGACQPCPINCTHSCYVDL 637  
 Db 1 GP--DHCKCAHFIDGPHCAVCACGAVLGENDTL-VKRYADANAVCOLCHPCTCRGCKP 57  
 638 DDGCPAQRASPLTSIVSAVV-GILLVVLGVGVGILIKRQOKIRKYTRMRLLQETLE 696  
 Db 58 GLECP---NSKTPSIAGVGVGLCLLVVGLIGLYLRRR-HIVRKTRRLLOEHEL 113  
 697 VEPLTPSGAMPNOAOMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKI PVAIKVLR 756  
 Db 114 VEPLTPSGAMPNOAOMRILKETELRKVKVLSGAFGTYYKGLWIPGEKVKI PVAIKELR 173  
 757 ENTSPKANKEILDEAYVAGVGSFYVSRILGICLTSTVQLVTQLMPYGCCLLDHVRNRR 816  
 Db 174 EATSPKANKEILDEAYVAGVGSFYVSRILGICLTSTVQLVTQLMPYGCCLLDYIREHKN 233  
 817 LGSODLWCMQIAKGSYLEVDRLVHRDLAARVNLVYKSPNHVITDGLARLLDIDETE 876  
 Db 234 IGSOYLWCMQIAKGMNLYLEERLVLHRDLAARVNLVYKTPQHVKITDGLAKLGADEKE 293  
 877 YHADGKVPIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLE 936  
 Db 294 YHADGKVPIMMALESILHRITHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLE 353  
 937 KGERLPQPPICITIDVWIMVCMWIDSECRPRELVESEFMRMDPQRFVVIQ-NEULG 995  
 Db 354 KGERLPQPPICITIDVWIMVCMWIDSECRPRELVESEFMRMDPQRFVVIQ-NEULG 413  
 996 PASPLDSTFYSLLEDDDMGDLVDAEYLVPQGGFCDDPAPGAGMWHHRSSSTNSG 1055  
 Db 414 LPSPTDSKFYRTLMEEEDMEIYVDAEYLVPHQGF-----NSPST----- 454  
 1056 GGDLTGLGPESEEARSPPL-----APSEGAGSDVFDGLGMAKGLQSLPTHPSPPLQ 1110  
 Db 455 -----SRTPLLSSLSATSNNSATNCIDRNG-----H----- 481  
 1111 RYSEDPTVLPSETDGYAAPLTCSPQPEYVNOQDVRRPQPSREGPLPAARAGAT-LE 1169  
 Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMQNGIYNFISLTAISK 523  
 1170 AKTISPGKGVKDVPAFAGAVENPEYL 1197  
 Db 524 LPIDSRYN-----SHSTAVDNPYL 544

RESULT 14

S00727  
 C:Species: avian erythroblastosis virus  
 C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 28-Feb-1997  
 C:Accession: S00727  
 R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.  
 Oncogene Res. 1, 265-278, 1987  
 A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mut  
 A:Reference number: S00727; MUID:88217326; PMID:2897102  
 A:Accession: S00727  
 A:Molecule type: DNA  
 A:Residues: 1-545 <SCO>  
 A:Cross-references: EMBL.X06943  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; phosphotransferase  
 F.135-400/Domain: protein kinase homology <KIN>  
 F.143-151/Region: protein kinase ATP-binding motif

Query Match 24.2%; Score 1640; DB 2; Length 545;  
 Best Local Similarity 54.9%; Pred. No. 2.6e-63;

Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

```

Qy 578 GPEADQVCACAHYKDPPECVACRPSGVKPDLSYMPIMKFPDEBGAQCPPCINCTHSCVDLDDKCCPAEQ 637
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GP--DHGKCAHFIIDPHCVACAPAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCGPGLCEGCP-- 57
Qy 638 DDKCCPAEQRASPILTSVSNV-GILVVVLGVGVGLIKRQOKIRKYMRLLOETLEVERLTPSGA 696
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 GLECCP--NGSKTPSIAAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRLLORELVEPLTPSGE 113
Qy 697 VEPILTPSGAMPNOAQMRILKETELRKVKVLSGAFGVYKGIWIPDENVKIPVAIVLRENTSPKAN 756
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 114 VEPILTPSGEAPNOAHRLKETEFKVKVLSGAFGVYKGIWIPBEKKTIPVAIKELRENTSPKAN 173
Qy 757 ENTSPKANKEILDEAYVMAVGSPPVSRLLGICLTSTVOLVQLMPYGCCLLDHVRNRRGLSGD 816
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 174 EATSPKANKEILDEAYVMAVSDNPHVCRLGICLTSTVOLITQLMPYGCCLLDYIREHKDN 233
Qy 817 LGSODLLNMCQIACKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDPEGLARLLDIDETEYHAC 876
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 IGSQYLLNMCVQIAKGMNLYEERHLVHRDLAARNVLYKTPQHVKITDPEGLAKOLGADKEKE 293
Qy 877 YHADGCKVPIKMMALLESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 294 YHAEKGVPPIKMMALLESILRRITTHOSDWSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
Qy 937 KGERLPOPICTIDVYMIWKCMMIDSECRPRFRELVSFSRMAADPQREVVIO-NEDLG 995
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 354 KGERLPOPICTIDVYMIWKCMMIDSECRPRFRELVSFSRMAADPQREVVIOVIGDERMH 413
Qy 996 PASLDDSTFYSRLLEDMDGLVDABEYLVPOQGFPCPDPAFGAGMHHRRSSSTRSG 1055
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 LPSTJDSKFPRTLMEEEDMEDIVDADEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDITLGLERSEEARSPPL-----APSEGAGSDVFPDGLGMAKGLQSLPTHDPSPLO 1110
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 481
Qy 1111 RYSEDPTVLPSETDGVVAELTCSPPQPEYVNPDPVRPQPSREGPLPAARPAAGAT-LER 1169
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKSTAMVQOIVYISLTAISK 523
Qy 1170 AKTLSPKNGVAKVFAFGAVENPEYL 1197
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 524 LPMDSRYON-----SHSTAVDNPEYL 544

```

## RESULT 15

B44776

Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)

C/Species: avian erythroblastosis virus

C/Date: 28-Apr-1993 #sequence\_revision 28-Apr-1993 #text\_change 04-Feb-2000

C/Accession: B44776

R/Buskin, A.; Jackson, J.; Bishop, J.M.; McCarty, D.J.; Schatzman, R.C.

Oncogene 5, 15-24, 1990

A/Title: Six amino acids from the retroviral gene gag greatly enhance the transforming p

A/Reference number: A44776; MUID:90206603; PMID:1969616

A/Accession: B44776

A/Status: Preliminary

A/Molecule type: mRNA

A/Residues: 1-540 <BRU>

A/Cross-references: GB:X52211

C/Genetics:

A/Gene: erbB

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

F/330-395/Domain: protein kinase homology <KIN>

F/138-146/Region: protein kinase ATP-binding motif

Query Match

23.9%; Score 1623; DB 2; Length 540;

Best Local Similarity 54.9%; Pred. No. 1,3e-62;

Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 14;

```

Qy 587 CAHYKDPPECVACRPSGVKPDLSYMPIMKFPDEBGAQCPPCINCTHSCVDLDDKCCPAEQ 646
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 CAHFIDGPHCVACAPAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCGPGLCEGCP-- 58
Qy 647 RASPLTSVSNV-GILVVVLGVGVGLIKRQOKIRKYMRLLOETLEVERLTPSGA 705
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 NGSKTPSIAAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRLLORELVEPLTPSGE 117
Qy 706 MPNOAQMRILKETELRKVKVLSGAFGVYKGIWIPDENVKIPVAIVLRENTSPKAN 765
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 APNOAHRLKETEFKVKVLSGAFGVYKGIWIPBEKKTIPVAIKELRENTSPKAN 177
Qy 766 EILDEAYVMAVGSPPVSRLLGICLTSTVOLVQLMPYGCCLLDHVRNRRGLSGD 825
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 EILDEAYVMAVSDNPHVCRLGICLTSTVOLITQLMPYGCCLLDYIREHKDNIGSQYLLNM 237
Qy 826 CMQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDPEGLARLLDIDETEYHAC 885
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 CVQIAKGMNLYEERHLVHRDLAARNVLYKTPQHVKITDPEGLAKOLGADKEKEYHAEKGVP 297
Qy 886 IKMMALLESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPP 945
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 298 IKMMALLESILRRITTHOSDWSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPOPP 357
Qy 946 ICTIDVYMIWKCMMIDSECRPRFRELVSFSRMAADPQREVVIO-NEDLG 1004
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 ICTIDVYMIWKCMMIDSECRPRFRELVSFSRMAADPQREVVIOVIGDERMH 417
Qy 1005 YRSLLEDMDGLVDABEYLVPOQGFPCPDPAFGAGMHHRRSSSTRSG 1064
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 418 YRSLLEDMDMEDIVDADEYLVPHQGF-----NSPST--- 449
Qy 1065 PSEEARSPPL-----APSEGAGSDVFPDGLGMAKGLQSLPTHDPSPLO 1119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 450 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 476
Qy 1120 LPSETDGVVAELTCSPPQPEYVNPDPVRPQPSREGPLPAARPAAGAT-LERAKTLSPGN 1178
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 477 -PVREDGFL-----PAPEYVNO--LMPKKSTAMVQOIVYISLTAISK 527
Qy 1179 GVVDVFAFGAVENPEYL 1197
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 528 -----SHSTAVDNPEYL 539

```

Search completed: July 22, 2003, 09:08:51

Job time : 29.9062 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds  
(without alignments)  
5088.033 Million cell updates/sec

Title: SEQ4-210-224-12

Perfect score: 6789  
Sequence: 1 MELALCMGGLLALLPPGA.....TFKGTPTANPEYIGLDVVP 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6698	98.7	1255	1	ERB2_HUMAN
2	5886	86.7	1257	1	ERB2_RAT
3	5876.5	86.6	1234	1	ERB2_MESAU
4	3074	45.3	1210	1	EGFR_HUMAN
5	3053	45.0	1210	1	EGFR_MOUSE
6	2909.5	42.9	1308	1	ERB4_HUMAN
7	2890	38.7	1308	1	ERB4_RAT
8	2625.5	34.7	1167	1	XMRK_XIPMA
9	2354.5	34.7	1167	1	ERB3_HUMAN
10	2283.5	33.6	1339	1	ERB3_RAT
11	1891	27.9	1426	1	EGFR_DROME
12	1749.5	25.8	634	1	ERBB_ALV
13	1703	25.1	604	1	ERBB_AVIER
14	1630	24.0	540	1	ERBB_AVIEU
15	1532	22.6	703	1	EGFR_CHICK
16	1255	18.5	1323	1	LT23_CAEL
17	1142.5	16.8	245	1	ERB2_MOUSE
18	675	9.9	1363	1	ILPR_BRALA
19	673	9.9	1382	1	INSR_HUMAN
20	666	9.8	1383	1	INSR_RAT
21	665.5	9.8	1372	1	INSR_MOUSE
22	665	9.8	1297	1	IRR_HUMAN
23	662	9.8	1300	1	IRR_MOUSE
24	662	9.8	1607	1	MIRP_LYMST
25	659.5	9.7	1300	1	IRR_CAVPO
26	642	9.5	1477	1	HTK7_HYDAT
27	603	8.9	1367	1	IGIR_HUMAN
28	596	8.8	1373	1	IGIR_MOUSE
29	595	8.8	987	1	EPB4_HUMAN
30	592.5	8.7	1370	1	IGIR_RAT
31	589.5	8.7	977	1	EPB2_MOUSE
32	589	8.7	987	1	EPB4_MOUSE
33	588	8.7	1114	1	RET_HUMAN

34	588	8.7	1390	1	INSR_AEDAE
35	588	8.7	2146	1	INSR_DROME
36	584.5	8.6	976	1	EPB2_HUMAN
37	582.5	8.6	984	1	EPB1_RAT
38	580.5	8.6	984	1	EPB1_CHICK
39	576.5	8.5	984	1	EPB1_HUMAN
40	573.5	8.4	1053	1	FAK1_CHICK
41	569	8.4	1068	1	FAK1_XENLA
42	567	8.4	757	1	HT16_HYDAT
43	567	8.4	902	1	EPB8_XENLA
44	566.5	8.3	985	1	EPB4_XENLA
45	563	8.3	1052	1	FAK1_MOUSE

## ALIGNMENTS

```

RESULT 1
ID      ERB2_HUMAN      STANDARD:      PRT: 1255 AA.
AC      P04626;
DT      13-AUG-1987 (Rel. 05, Created)
DT      13-AUG-1987 (Rel. 05, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE      (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE      surface receptor HER2) (MLN 19).
CN      ERB2 OR HER2 OR NGL OR NEU.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=86118653; Pubmed=3003577;
RA      Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA      Saito T., Toyoshima K.;
RT      "Similarity of protein encoded by the human c-erb-B-2 gene to
RT      epidermal growth factor receptor.";
RL      Nature 319:230-234(1986).
[2]
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=86070181; Pubmed=2999974;
RA      Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA      McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA      Francke U., Levinson A., Ullrich A.;
RT      "Tyrosine kinase receptor with extensive homology to EGF receptor
RT      shares chromosomal location with neu oncogene.";
RL      Science 230:1132-1139(1985).
[3]
RN      [3]
RP      SEQUENCE OF 717-1031 FROM N.A.
RX      MEDLINE=86016729; Pubmed=2995967;
RA      Semba K., Kanata N., Toyoshima K., Yamamoto T.;
RT      "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT      c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT      human salivary gland adenocarcinoma.";
RL      Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
[4]
RN      [4]
RP      VARIANTS VAL-654 AND VAL-655.
RX      MEDLINE=93194196; Pubmed=8095488;
RA      Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT      "Characterization of a new allele of the human ERBB2 gene by allele-
RT      specific competition hybridization.";
RL      Genomics 15:426-429(1993).
-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
ALPHA AND AMPHIREGULIN.
-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
-1- SUBCELLULAR LOCATION: Type I membrane protein.

```

FT	DISULFID	563	576	BY SIMILARITY.
FT <th>DISULFID</th> <td>567</td> <td>584</td> <td>BY SIMILARITY.</td>	DISULFID	567	584	BY SIMILARITY.
FT <th>DISULFID</th> <td>587</td> <td>596</td> <td>BY SIMILARITY.</td>	DISULFID	587	596	BY SIMILARITY.
FT <th>DISULFID</th> <td>600</td> <td>623</td> <td>BY SIMILARITY.</td>	DISULFID	600	623	BY SIMILARITY.
FT <th>DISULFID</th> <td>626</td> <td>634</td> <td>BY SIMILARITY.</td>	DISULFID	626	634	BY SIMILARITY.
FT <th>DISULFID</th> <td>630</td> <td>642</td> <td>BY SIMILARITY.</td>	DISULFID	630	642	BY SIMILARITY.
FT <th>MOD. RES</th> <td>1139</td> <td>1139</td> <td>PHOSPHORYLATION (AUTO-) (BY SIMILARITY).</td>	MOD. RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT <th>MOD. RES</th> <td>1248</td> <td>1248</td> <td>PHOSPHORYLATION (AUTO-) (BY SIMILARITY).</td>	MOD. RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT <th>CARBOHYD</th> <td>68</td> <td>68</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>124</td> <td>124</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	124	124	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>187</td> <td>187</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	187	187	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>259</td> <td>259</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	259	259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>530</td> <td>530</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>571</td> <td>571</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	571	571	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>629</td> <td>629</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <th>VARIANT</th> <td>654</td> <td>654</td> <td>I -&gt; V.</td>	VARIANT	654	654	I -> V.
FT <th>VARIANT</th> <td>655</td> <td>655</td> <td>/FTID=VAR_004077.</td>	VARIANT	655	655	/FTID=VAR_004077.
FT <th>VARIANT</th> <td>655</td> <td>655</td> <td>I -&gt; V.</td>	VARIANT	655	655	I -> V.
FT <th>VARIANT</th> <td>655</td> <td>655</td> <td>/FTID=VAR_004078.</td>	VARIANT	655	655	/FTID=VAR_004078.
FT <th>CONFLICT</th> <td>1170</td> <td>1170</td> <td>P -&gt; A (1N REF. 2).</td>	CONFLICT	1170	1170	P -> A (1N REF. 2).
FT <th>SEQUENCE</th> <td>1255 AA; 137909 MM; 39E9FD6A04DC6962 CRC64;</td> <td></td> <td></td>	SEQUENCE	1255 AA; 137909 MM; 39E9FD6A04DC6962 CRC64;		
Query Match				
Best Local Similarity 98.7%; Score 6698; DB 1; Length 1255;				
Matches 1240; Conservative 98.8%; Pred. No. 0; Mismatches 14; Indels 0; Gaps 0				
QY	1	MEALALCWMGLLALLPRGAASVCTGTDMRLRLPASETHLDMRLHYOGCGVQGNL	60	
DB	1	MEALALCWMGLLALLPRGAASVCTGTDMRLRLPASETHLDMRLHYOGCGVQGNL	60	
QY	61	ELTYLPTNASLSFLODIOEVQGYVLIANHQVQVPLQRLRIYRGTOLFEDNYALAVDNG	120	
DB	61	ELTYLPTNASLSFLODIOEVQGYVLIANHQVQVPLQRLRIYRGTOLFEDNYALAVDNG	120	
QY	121	DPLNNTTPEVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA	180	
DB	121	DPLNNTTPEVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA	180	
QY	181	LTLIDTNSRAHPSPSPCKSGKSGCESEDDOSLTRYCAGGACAKGGLPOLYIKANSK	240	
DB	181	LTLIDTNSRAHPSPSPCKSGKSGCESEDDOSLTRYCAGGACAKGGLPOLYIKANSK	240	
QY	241	FIGITELKRSDDLACLFHNSGICELHPCALVTYNTDFESMNPREGRYTFGASCVTACP	300	
DB	241	AAGCTGPRHSDCLACLFHNSGICELHPCALVTYNTDFESMNPREGRYTFGASCVTACP	300	
QY	301	YNYLSTDVGSCTLVCPRLHNOEVTABDGTORCEKSPCARVCYGLCMEHLREYRAVTSAN	360	
DB	301	YNYLSTDVGSCTLVCPRLHNOEVTABDGTORCEKSPCARVCYGLCMEHLREYRAVTSAN	360	
QY	361	IOEPFAGCKKIFGSLAFELPESFGDPDASNPAPLOPELOLVFETLEETIGVLYISAMPDLSL	420	
DB	361	IOEPFAGCKKIFGSLAFELPESFGDPDASNPAPLOPELOLVFETLEETIGVLYISAMPDLSL	420	
QY	421	DLVSFQNLQVIRGRILIHNGAYSLLTQGLGISWLGRLSRLRELSGALLIHNHTLCEVHTV	480	
DB	421	DLVSFQNLQVIRGRILIHNGAYSLLTQGLGISWLGRLSRLRELSGALLIHNHTLCEVHTV	480	
QY	481	PMDOLFRRPHQALLHTANRPDECEYGEGLACHQLCARHGCMWGPPIQCVNCSQFLRGQEC	540	
DB	481	PMDOLFRRPHQALLHTANRPDECEYGEGLACHQLCARHGCMWGPPIQCVNCSQFLRGQEC	540	
QY	541	VEECQVTLQGLPREYVNAHRLCPHCEQCPQNSVTCFGEADQCVCVCAHYKDPFCVARC	600	
DB	541	VEECQVTLQGLPREYVNAHRLCPHCEQCPQNSVTCFGEADQCVCVCAHYKDPFCVARC	600	
QY	601	PSGKVPDLSTYMPIMKFPDPEBEGACQPCPINCTHSQVLDLDKCGPABEGRASPLTSISAVVG	660	
DB	601	PSGKVPDLSTYMPIMKFPDPEBEGACQPCPINCTHSQVLDLDKCGPABEGRASPLTSISAVVG	660	
QY	661	ILLVVLGVNFGVILLKRRQOKLRKXTMRLLOETELVEPLTPSGAMPNQAOMRLKETEL	720	
DB	661	ILLVVLGVNFGVILLKRRQOKLRKXTMRLLOETELVEPLTPSGAMPNQAOMRLKETEL	720	

QY	721	RRVKVLGSGACTYVKKGIWIPGENVKTPVAIKVLRENTSPRANKKEILIDENYVWAGVSP	780
Db	721	RRVKVLGSGAECTYVKKGIWIPGENVKTPVAIKVLRENTSPRANKKEILIDENYVWAGVSP	780
QY	781	YVSRLLGICLSTVOLVTLQMLPFGCLLDHVENRGRGLSODLLNMCQIAAGMSYLEDDR	840
Db	781	YVSRLLGICLSTVOLVTLQMLPFGCLLDHVENRGRGLSODLLNMCQIAAGMSYLEDDR	840
QY	841	LVHRDLAARNLVKSPNHYKTTDFGLARLLDIDETEHADGKVPKIMMALESILRRFT	900
Db	841	LVHRDLAARNLVKSPNHYKTTDFGLARLLDIDETEHADGKVPKIMMALESILRRFT	900
QY	901	HOSDWSVGVYWEMLTFGAKPYDGIIPAREIPDLLEKEERLPPOPCTIDVYMTVMCM	960
Db	901	HOSDWSVGVYWEMLTFGAKPYDGIIPAREIPDLLEKEERLPPOPCTIDVYMTVMCM	960
QY	961	IDSECRPPRRELVSFESRMDRPPORFVVIIONEDLGAPSLDSTFYRSLLIEDMDGLVDA	1020
Db	961	IDSECRPPRRELVSFESRMDRPPORFVVIIONEDLGAPSLDSTFYRSLLIEDMDGLVDA	1020
QY	1021	EELYVPOGFFCCPPDPAPAGMWHHRSSSTRSGGDLTLGLEBSEEARSPPLADSEG	1080
Db	1021	EELYVPOGFFCCPPDPAPAGMWHHRSSSTRSGGDLTLGLEBSEEARSPPLADSEG	1080
QY	1081	AGSDVFDDDLGGAKKGLQSLPTHPSPFLORSDEPTVPLSETDGYVAPLTCSPOPEYV	1140
Db	1081	AGSDVFDDDLGGAKKGLQSLPTHPSPFLORSDEPTVPLSETDGYVAPLTCSPOPEYV	1140
QY	1141	NOPDVRPOPSPBREPPLPAARPAGATLBERAKTISPGRNVYKDVAFAGAVENPEYLTPO	1200
Db	1141	NOPDVRPOPSPBREPPLPAARPAGATLBERAKTISPGRNVYKDVAFAGAVENPEYLTPO	1200
QY	1201	GGAAPQHPPPAPFSAFNLVYWDODDPPERGAPSTFFGTPTAENPEYLGIDVEV	1255
Db	1201	GGAAPQHPPPAPFSAFNLVYWDODDPPERGAPSTFFGTPTAENPEYLGIDVEV	1255
RESULT 2			
ERB2_RAT			
ID_ERB2_RAT	STANDARD;	PRT; 1257 AA.	
AC	P06494;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).		
GN	ERBB2 OR NEU.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_Taxid:10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Neuroblastoma;		
RX	MEDLINE=86118662; PubMed=3945311;		
RA	Bargmann C.I., Hung M.-C., Weinberg R.A.;		
RT	"The neu oncogene encodes an epidermal growth factor receptor-related		
RT	protein.";		
RL	Nature 319:226-230(1986).		
RL	[2]		
RP	SEQUENCE OF 852-905 FROM N.A.		
RC	TISSUE=Sciatic nerve;		
RX	MEDLINE=91222560; PubMed=2025425;		
RA	Lai C., Lemke G.;		
RT	"An extended family of protein-tyrosine kinase genes differentially		
RT	expressed in the vertebrate nervous system.";		
RL	Neuron 6:691-704(1991).		
RN	[3]		
RP	STRUCTURE BY NMR OF 650-668		
RX	MEDLINE=92151581; PubMed=1346763;		
RA	Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,		

```

RA Newman R; Cumpston M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RL oncogenic and oncogenic forms of the new protein.";
EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREULINS DO NOT INTERACT WITH IT ALONE. GP10 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X03362; CAAT27059.1; ALT_INIT.
CC PIR; A24562; TVRTNU.
CC DR HSSP; P11362; 1FGK.
CC DR InterPro; IPR000494; EGFR_L_domain.
CC DR InterPro; IPR000719; Euk_Pkinase.
CC DR InterPro; IPR002174; Furin-like.
CC DR InterPro; IPR001245; Tyr_Pkinase.
CC DR InterPro; IPR004019; YLP motif.
CC DR Pfam; PF000069; Pkinase; 1.
CC DR Pfam; PF00757; Furin-like; 1.
CC DR Pfam; PF01030; Recep_L_domain; 2.
CC DR Pfam; PF02751; YLP_2.
CC DR ProDom; PD000001; Euk_pkinase; 1.
CC DR SMART; SM00261; FU; 3.
CC DR SMART; SM00219; TYRC; 1.
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC KW Proto-oncogene; Disease mutation.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
CC FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
CC FT TRANSSEM 655 677 POTENTIAL.
CC FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 159 369 CYS-RICH.
CC FT DOMAIN 473 646 CYS-RICH.
CC FT DOMAIN 722 989 CYS-RICH.
CC FT NP_BIND 728 736 PROTEIN_KINASE.
CC FT BINDING 755 755 ATP (BY SIMILARITY).
CC FT ACT_SITE 847 847 ATP (BY SIMILARITY).
CC FT DISULFID 196 205 BY SIMILARITY.
CC FT DISULFID 200 213 BY SIMILARITY.
CC FT DISULFID 221 228 BY SIMILARITY.
CC FT DISULFID 225 236 BY SIMILARITY.
CC FT DISULFID 241 253 BY SIMILARITY.
CC FT DISULFID 253 265 BY SIMILARITY.
CC FT DISULFID 256 265 BY SIMILARITY.
CC FT DISULFID 269 296 BY SIMILARITY.
CC FT DISULFID 300 312 BY SIMILARITY.
CC FT DISULFID 316 332 BY SIMILARITY.
CC FT DISULFID 335 339 BY SIMILARITY.
CC FT DISULFID 513 522 BY SIMILARITY.
CC FT DISULFID 517 530 BY SIMILARITY.
CC FT DISULFID 533 542 BY SIMILARITY.
CC FT DISULFID 546 562 BY SIMILARITY.
CC FT DISULFID 565 578 BY SIMILARITY.

```



[illegible]

Db	781	SPVSRLLGICLSTVQTLVQMLPFGCLLDHVRHGRGLSGODLLNCVOJLAKGMSYLED	840
QY	839	VRLYHRPLAARNVLYKSPNHVKITDFGLARLLDIDETERYHADGGKVPILKMALESILRR	898
Db	841	VRLYHRPLAARNVLYKSPNHVKITDFGLARLLDIDETERYHADGGKVPILKMALESILRR	900
QY	899	FTTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICITIDVYIMVVC	958
Db	901	FTTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICITIDVYIMVVC	960
QY	959	WMIDSECRPRELVESEFSRMAADPQRVUVIONELGPASPIDSTFFRSLLEDDDDMGDLV	1018
Db	961	WMIDSECRPRELVESEFSRMAADPQRVUVIONELGPASPIDSTFFRSLLEDDDDMGDLV	1020
QY	1019	DAEELVYVQGGFFCCDDPAPAGAGMYHHRSSSTFGGDLTLGLEPSEEARSPPLAPS	1078
Db	1021	DAEELVYVQGGFFCCDDPAPAGAGMYHHRSSSTFGGDLTLGLEPSEEARSPPLAPS	1080
QY	1079	EGAGSDVDFEGDGLGMGAOKGLQSLPTHDPSPLORYSEDPTVPLPSFNDGVYAPLTCSPQPE	1138
Db	1081	EGAGSDVDFEGDGLGMGAOKGLQSLPTHDPSPLORYSEDPTVPLPSFNDGVYAPLTCSPQPE	1140
QY	1139	YVNPDPVRFQPPSPRGGPLPAPAPACATLBERATLSLPGKNGVYKDVAFAGAVENPEYLT	1198
Db	1141	YVNPDPVRFQPPSPRGGPLPAPAPACATLBERATLSLPGKNGVYKDVAFAGAVENPEYLT	1200
QY	1199	PGCGAPOPHPAPFAPFNDLYYMDODPERGAPSPSTFKGTPTAENPEYLGLDVY	1255
Db	1201	PREGTASPPHSPAFSPAPFNDLYYMDODPERGAPSPSTFKGTPTAENPEYLGLDVY	1257
RESULT 3			
ERB2_MESAU	STANDARD;	PRT;	1254 AA.
ID	ERB2_MESAU		
AC	Q60553;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase <i>erbB-2</i> precursor (BC 2.7.1.112)		
DN	(p185 <sup>erbB2</sup> ) (NEU proto-oncogene) (C-erbB-2).		
OS	ERBB2 OR NEU.		
OS	Mesocricetus auratus (Golden hamster).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;		
OC	Mesocricetus.		
OK	NCBI_TaxID=10036;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Nerve;		
RX	MEDLINE=94193007; PubMed=7908275;		
RA	Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,		
RA	Yamazaki Y., Ishikawa T.;		
RT	"Cloning and activation of the Syrian hamster neu proto-oncogene.";		
RL	Gene 140:251-255(1994).		
CC	-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,		
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A		
CC	POTENTIAL LIGAND FOR THIS RECEPTOR, NOT ACTIVATED BY EGF, TGF-		
CC	ALPHA AND AMPHIREGULIN (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS		
CC	(POTENTIAL).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE		
CC	RESIDUES.		
CC	-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: D16295; BAA03801.1; .

DR HSSP: P1362; 1FGK.

DR InterPro: IPR000494; EGFR\_L\_domain.

DR InterPro: IPR000719; Euk\_kinase.

DR InterPro: IPR002174; Furin-like.

DR InterPro: IPR001245; Tyr\_kinase.

DR InterPro: IPR004019; YLP\_motif.

DR Pfam: PF00069; kinase; 1.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF01030; Recept\_L\_domain; 2.

DR Pfam: PF02757; YLP; 2.

DR ProDom: PD000001; Euk\_kinase; 1.

DR SMART: SM00261; FU; 3.

DR SMART: SM00219; Tyrc; 1.

DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.

DR PROSITE: PS0011; PROTEIN\_KINASE\_DOM; 1.

DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

KW Proto-oncogene; Disease mutation.

FT SIGNAL 1 21

FT CHAIN 22 1254

FT DOMAIN 22 652

FT TRANSMEM 653 675

FT DOMAIN 676 1254

FT DOMAIN 158 368

FT DOMAIN 472 644

FT DOMAIN 720 987

FT NP\_BIND 726 734

FT BINDING 753 753

FT ACT\_SITE 845 845

FT DISULFID 195 204

FT DISULFID 199 212

FT DISULFID 236 244

FT DISULFID 240 252

FT DISULFID 255 264

FT DISULFID 268 295

FT DISULFID 299 311

FT DISULFID 315 331

FT DISULFID 334 338

FT DISULFID 511 520

FT DISULFID 515 528

FT DISULFID 531 540

FT DISULFID 544 560

FT DISULFID 563 576

FT DISULFID 567 584

FT DISULFID 587 596

FT DISULFID 600 623

FT DISULFID 626 634

FT DISULFID 630 642

FT MOD\_RES 1139 1139

FT MOD\_RES 1247 1247

FT CARBOHYD 68 68

FT CARBOHYD 125 125

FT CARBOHYD 187 187

FT CARBOHYD 259 259

FT CARBOHYD 530 530

FT CARBOHYD 571 571

FT CARBOHYD 629 629

FT CARBOHYD 658 658

FT VARIANT 659 659

FT VARIANT 659 659

SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 86.6%; Score 5876.5; DB 1; Length 1254;

Best Local Similarity 86.5%; Pred. No. 1,1e-310;

Matches 1086; Conservative 58; Mismatches 110; Indels 1; Gaps 1;

QY 1 MELALATGRWGILLALLPFGAASGVCTGTDKRLRLPASPETHLMLRLHLYGGCCVVGQNL 60

DB 1 MELAAWGMGLLALLSLSGASGVCTGTDKRLRLPASPETHLMLRLHLYGGCCVVGQNL 60

QY 61 ELTYLPTNASTSLFDIDIQEVGVYLAHQVQVPLQRLRIYRGTLQFEDNALAVLDNG 120

DB 61 ELTYLPANATSLFDIDIQEVGVYLAHQVQVPLQRLRIYRGTLQFEDKVALAVLDNR 120

QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNQLA 180

DB 121 DPLDNVTTATGRTPGELRELQRLSLTEILKGVLIQRNPOLCYODTIVLWKDVFRKNQQLA 180

QY 181 LTLIDNRSRACHPCSPMKCKSRCKGSESEDDOSLTRCYAGCARGCKRPLQYITANSK 240

DB 181 PVDIDNRSRACHPCSPMKCKSRCKGSESEDDOSLTRCYAGCARGCKRPLQYITANSK 240

QY 241 FIGITELKHSDDLACILFHNHSGICELHCPALVTYNTDTFESNPNPEGRYTFGASCTYAP 300

DB 241 AAGCTGPRHSDCLACILFHNHSGICELHCPALVTYNTDTFESNPNPEGRYTFGASCTYAP 300

QY 301 YNYLSTDVGSCTLVCPRLHQEVTAEDGTQRCCKSKPCARVCYGLQMEHLREVAVTSAN 360

DB 301 YNYLSTDVGSCTLVCPRLHQEVTAEDGTQRCCKSKPCARVCYGLQMEHLREVAVTSAN 360

QY 361 IOEFAGCKKIRGSLAFLESFPGDPAASNTAPLQPELOVPELLEITGLYISAMPDSL 420

DB 361 IOEFAGCKKIRGSLAFLESFPGDPAASNTAPLQPELOVPELLEITGLYISAMPDSL 420

QY 421 DLSTFQNLQVIRGRILHNQAVSLTLOGLGISWGLRSLRELSSGLALIHNNHLCFVHTV 480

DB 421 DLSTFQNLQVIRGRILHNQAVSLTLOGLGISWGLRSLRELSSGLALIHNNHLCFVHTV 480

QY 481 PWDQLEFRNPQALHTANRPEDECVGEGLACHQLCARHGMCPGTQCVNCSQFLRGQEC 540

DB 481 PWDQLEFRNPQALHTANRPEDECVGEGLACHQLCARHGMCPGTQCVNCSQFLRGQEC 540

QY 541 VEECVQGLQPREYNAHCLPCHCECQPNQSVTCFPEADQCAACAHKDPFPCVAC 600

DB 541 VEECVQGLQPREYNAHCLPCHCECQPNQSVTCFPEADQCAACAHKDPFPCVAC 600

QY 601 PSQVPRDLSYMPIMKFPDEBEGACOPCINCHSCVDLDDKGPAPORASPLSISAVVG 660

DB 601 PSQVPRDLSYMPIMKFPDEBEGACOPCINCHSCVDLDDKGPAPORASPLSISAVVG 660

QY 661 ILLVVLGVFGLIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKETEL 720

DB 661 ILLVVLGVFGLIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKETEL 720

QY 721 RKVRLTSGAGTGYVKGIMIDGENVKIPVAIKVLRNTSPRANKELIDEAVYMGVSGP 780

DB 721 RKVRLTSGAGTGYVKGIMIDGENVKIPVAIKVLRNTSPRANKELIDEAVYMGVSGP 780

QY 781 YVSRLLGICLSTVOLVQLMPYGLDHPVENRGLSODLLNMCQIAKMSYLEDDR 840

DB 781 YVSRLLGICLSTVOLVQLMPYGLDHPVENRGLSODLLNMCQIAKMSYLEDDR 840

QY 841 LVHRLDARNAVLYVSPNFKITTDGLARLLDIDETEHADGKVIKIMIALESILRRRT 900

DB 841 LVHRLDARNAVLYVSPNFKITTDGLARLLDIDETEHADGKVIKIMIALESILRRRT 900

QY 901 HQSDVMSGVTVWELMTFGARPYDIPAREIPDLLEKEKRLPQPICTIDVYMIWKKCM 960

DB 901 HQSDVMSGVTVWELMTFGARPYDIPAREIPDLLEKEKRLPQPICTIDVYMIWKKCM 960

QY 961 IDSECRPRFRLVSEFSRMRDPPQFVYIQNEDJGAPASPLSTFRSLLEDMDMDLVDA 1020

DB 961 IDSECRPRFRLVSEFSRMRDPPQFVYIQNEDJGAPASPLSTFRSLLEDMDMDLVDA 1020

QY 1021 EBYLVPOGFPDPAAGGMVHRHRSSTRSGGDLTLGLSESEAPRSPPLAPSEEG 1080

DB 1021 EBYLVPOGFPDPAAGGMVHRHRSSTRSGGDLTLGLSESEAPRSPPLAPSEEG 1080

QY 1081 AGSDVFPQDLAGMAKGLQSLPTHDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

DB 1081 AGSDVFPQDLAGMAKGLQSLPTHDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy	1141	NOPVTRPPSPRGSGPLPPAAPAGCTTLERATLSPGKNGVVKVDPAFCGAENPEYLTPO	1200
Db	1141	NQPERPPPLTTPCEPLPPVPAPAGATTLERPTLSPGKNGVKVDFTFGAAVENEPEYLVPR	1200
Qy	1201	CGAAPQHPHPAPFASPAFDNLTYWQDDPERGAPSPSTFGKTPTAANPEYLTGLDVEV	1255
Db	1201	CGSASQPH-PPALCGAFDNLTYWQDDPERGSPSTFTFGTPTAENPEYLTGLDVEV	1254
RESULT 4			
ID	EGFR_HUMAN	STANDARD; PRT; 1210 AA.	
AC	P00533; P06268; Q14225; Q9UMD7; Q9UM08; Q9UMG5; Q92795; Q00732;		
CC	O00686; Q9B252; Q9H2C9; Q9GZX1; Q9H3C9;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor		
DE	protein-tyrosine kinase ErbB-1).		
GN	EGFR OR ERBB1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RN	SEQUENCE FROM N.A. (ISOFORM 1).		
RX	MEDLINE=84219729; PubMed=6328312;		
RA	Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,		
RA	Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,		
RA	Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;		
RT	"Human epidermal growth factor receptor cDNA sequence and aberrant		
RT	expression of the amplified gene in A431 epidermoid carcinoma cells.";		
RL	Nature 309:418-425 (1984).		
RN	[2]		
RN	SEQUENCE FROM N.A. (ISOFORM 2).		
RC	TISSUE=Placenta;		
RX	MEDLINE=9538257; PubMed=7654368;		
RA	Ilekta J.V., Stark B.C., Scoccia B.;		
RT	"Possible role of variant RNA transcripts in the regulation of		
RT	epidermal growth factor receptor expression in human placenta.";		
RL	Mol. Reprod. Dev. 41:149-156(1995).		
RN	[3]		
RN	SEQUENCE FROM N.A. (ISOFORM 2).		
RC	TISSUE=Placenta;		
RX	MEDLINE=97078686; PubMed=8918811;		
RA	Reiter J.L., Maible N.J.;		
RT	"A 1.8 kb alternative transcript from the human epidermal growth		
RT	factor receptor gene encodes a truncated form of the receptor.";		
RL	Nucleic Acids Res. 24:4050-4056(1996).		
RN	[4]		
RN	SEQUENCE FROM N.A. (ISOFORM 2).		
RC	TISSUE=Placenta;		
RX	MEDLINE=97256547; PubMed=9103388;		
RA	Ilekta J.V., Gariti J., Niederberger C., Scoccia B.;		
RT	"Expression of a truncated epidermal growth factor receptor-like		
RT	protein (TRGFP) in ovarian cancer.";		
RL	Gynecol. Oncol. 65:36-41(1997).		
RN	[5]		
RN	SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).		
RC	TISSUE=Placenta;		
RX	MEDLINE=21100872; PubMed=11161793;		
RA	Reiter J.L., Thiradgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,		
RA	Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,		
RA	Balaubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,		
RA	Maible N.J.;		
RT	"Comparative genomic sequence analysis and isolation of human and		
RT	mouse alternative EGFR transcripts encoding truncated receptor		
RT	isoforms.";		
RL	Genomics 71:1-20(2001).		
RN	[6]		
RP	SEQUENCE OF 575-687 FROM N.A.		
RA	Reiter J.L., Thiradgill D.W., Danielson A.J., Schehl C.M.,		
RA	Lampland A.L., Balaubramaniam S., Crossley T.O., Magnuson T.R.,		

RA "Human and mouse alternative EGFR transcripts encoding only the  
RT extracellular domain of the receptor." ;  
RL Submitted (FEB-1999) to the EMBL/Genbank/DDBJ databases.  
RN [7]  
RP SEQUENCE OF 713-924 FROM N.A.  
RX MEDLINE=84196372; PubMed=6326561;  
RA Lin C.R., Chen W.S., Krueger W., Stolaraky L.S., Weber W.,  
RA Evans R.M., Verma I.M., Gali G.N., Rosenfield M.G.;  
RT "Expression cloning of human EGF receptor complementary DNA: gene  
RT amplification and three related messenger RNA products in A431  
RT cells.";  
RL Science 224:843-848(1984).  
RN [8]  
RP SEQUENCE OF 150-962 FROM N.A.  
RX MEDLINE=84245835; PubMed=6330563;  
RA Xu Y.H., Ishii S., Clark A.J.D., Sullivan M., Wilson R.K., Ma D.P.,  
RA Roe B.A., Merlino G.T., Pastan I.;  
RT "Human epidermal growth factor receptor cDNA is homologous to a  
RT variety of RNAs overproduced in A431 carcinoma cells.";  
RL Nature 309:806-810(1984).  
RN [9]  
RP SEQUENCE OF 1028-1210 FROM N.A.  
RX MEDLINE=85046483; PubMed=6093780;  
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,  
RA O'Malley B.W.;  
RT "Isolation of an evolutionarily conserved epidermal growth factor  
RT receptor cDNA from human A431 carcinoma cells.";  
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
RN [10]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=86821733; PubMed=3329716;  
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
RA Waterfield M.D.;  
RT "The human EGF receptor gene: structure of the 110 kb locus and  
RT identification of sequences regulating its transcription.";  
RL Oncogene Res. 1:375-396(1987).  
RN [11]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=91107677; PubMed=198448;  
RA Haley J.D., Waterfield M.D.;  
RT "Contributory effects of de novo transcription and premature  
RT transcript termination in the regulation of human epidermal growth  
RT factor receptor proto-oncogene RNA synthesis.";  
RL J. Biol. Chem. 266:1746-1753(1991).  
RN [12]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=85270438; PubMed=2991899;  
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
RT "Characterization and sequence of the promoter region of the human  
RT epidermal growth factor receptor gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
RN [13]  
RP SEQUENCE OF 540.  
RA Kohda D.;  
RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
RN [14]  
RP RECEPTOR ACTIVITY.  
RX MEDLINE=84191554; PubMed=6325948;  
RA Mroczkowski B., Mosig G., Cohen S.;  
RT "ATP-stimulated interaction between epidermal growth factor receptor  
RT and supercoiled DNA.";  
RL Nature 309:270-273(1984).  
RN [15]  
RP PHOSPHORYLATION.  
RX MEDLINE=89278137; PubMed=2543678;  
RA Marcolis B.L., Lax I., Kris R., Dombalagian M., Honneger A.M.,  
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
RT "All autophosphorylation sites of epidermal growth factor (EGF)  
RT receptor and HER/neu are located in their carboxyl-terminal tails."  
RL Identification of a novel site in EGF receptor.";  
RJ J. Biol. Chem. 264:10667-10671(1989).  
RN [16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RA MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts."  
 RL Growth Factors 13:121-132(1996).  
 RL [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-566 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668.  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor."  
 RL J. Biochem. 127:65-72(2000).  
 RL [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Oda M., Inagaki F., Iax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor."  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RL [19]  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens."  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X00588; CA25240.1; -  
 DR EMBL: U95089; AAB53063.1; -  
 DR EMBL: U48722; AAC50802.1; -  
 DR EMBL: U48723; AAC50804.1; -  
 DR EMBL: U48724; AAC50796.1; -  
 DR EMBL: U48725; AAC50797.1; -  
 DR EMBL: U48725; AAC50798.1; -  
 DR EMBL: U48725; AAC50798.1; -  
 Query Match 45.3%; Score 3074; DB 1; Length 1210;  
 Best Local Similarity 49.0%; Pred. No. 4.8e-159;  
 Matches 620; Conservative 176; Mismatches 363; Indels 106; Gaps 21;  
 QY 11 LILALPFGAA--STGYCTGTDMKRLRPASPTHLDMLRHLYGCGGVVQGNLELTLYPTN 68  
 DB 14 LLAALCPASALAEKVKCGGTSNKLTLQGTGFEDHFLSLQRMFNCEVVLGNLETTYVORN 73

QY ASLSEFDIOIEVOGVYLIAHNOVROVPLORLRIVGTOLFEEDNYALALVLDNGDPLNNTTP 128  
 DB YDLSELTQIEVAGVILALNLTVEIPLENDQIHGMNMYENSALVALVSTND----- 126  
 QY VTGASPGGLRELQSLSTELIKGVLIOENPOLCYODTILMKDIFHKNNQALTLTIDTR 188  
 DB ---ANKTLKELPMNRLQELIHGAVRFSNNPALCNVESIQMRDVISSDPLSMWSMDFOH 183  
 QY SRACIPGSPMKGSGSKWSESDQSLRTVYAGGCA-RCKGPLQYIKANSKFGITEL 247  
 DB LGSQCKCPSPNSGSCWAGBENCCYKTKIICAOQSCSRCKSPSDCCNCAAGCTGP 243  
 QY KHSQCLALHPNHSIGICELCPALVTYNTDFESPMPNEGRTFASCTVACPYYVLTSD 307  
 DB RESDVLVCRKFRDAETCTDCTCPPLMLYPTTYQMVNVEGKYSFATCYKKCPRYVYTD 303  
 QY VGSCTLVCPNLNQEVTAEVGTORECKSKPCARVCYGLMEHLREAVTSANIOEFAC 367  
 DB HGSQVRACGADSYEM-EDGVKCKKCGPCRKVCNGIIGIEFKDLSINATINIGHFNK 362  
 QY KKTIGSLAFLPESPDGPDASTAPLQEPQLOVFETLEITGYLYISAMPDSLPLDSVFON 427  
 DB TSISGDHLILPVAFRGDSFTHTPPLDPQELDLKTKVEITGFLLIQAMPENRTDLHAFEN 422  
 QY LOVIRGRILHNGAVSLTQGLGISWLGRLRELGSSALILHNHNLGFVHTVPDQFLR 487  
 DB LEIRGRKQGGQSLAVSLNLTSLGRLSEISDGVYIISGNKNLCTYANTINNKKLFG 482  
 QY NPHQALHTANRPDECEVGEGLACHQLCARGCHMGPGPTQCNCQSLFRGQCEVECRVL 547  
 DB TSGQKTKLISRGEMSCATQGVCHALCSPEGCMGPEPRDVCSCNNSRGRCVCKNLL 542  
 QY QGLREYVNAHCLPCHEPCOPNGSVTCFPEADQVCAAHYKDPFCVARGCPGVKPD 607  
 DB EGEPREFEYENSECICQCHPECLPQAMNITCTGRGPNICQCAHYIDGPHCVKTCFPAWGE 602  
 QY LSWPIMKFPDEBGACOPCPINCTHSCVDLDDKGCAPAGRASPLTISYSAVG---ILLY 664  
 DB NNTL-VMKYADAGVCHLCHPNCYCTGCPGLGCGPTNGPKIP-SIATGMGALLLLLV 659  
 QY VLVGVGILIKRROOKIKRYTMRRLQETELVEPLTPSGAMPNOQAKRIKETELRYK 724  
 DB VALGIG---LPMRRRIYRKRTLRLQERLVEPLTPSGAPNOLLRIKETEFKKIK 716  
 QY VLSGAGCTYVKGWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSPVYSR 784  
 DB VLSGAGCTYVKGWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSPVYSR 776  
 QY LIGICLSTVOLVQLMPYGLLDHVENRGRLSODLLNMCQIAKMSLTEDRLVHR 844  
 DB LIGICLSTVOLVQLMPYGLLDHVENRGRLSODLLNMCQIAKMSLTEDRLVHR 836  
 QY DLAAARNLVKSPNHVKTITDFGLARLIDIDETEHADGKVPKIMMALESIRRRTHOSD 904  
 DB DLAAARNLVKTPGVHKTITDFGLARLIDIDETEHADGKVPKIMMALESIRRRTHOSD 896  
 QY VMSYGVTVWELMTFGAPYDGIIPAREIPDLIEKGERLPQPPICITIDVYMIWVKCMIDSE 964  
 DB VMSYGVTVWELMTFGAPYDGIIPAREIPDLIEKGERLPQPPICITIDVYMIWVKCMIDSE 956  
 QY CRPRFRELVEEFMSMADPQRFVYIO-NEDLGPAASPLDSTYRSLDEDDMDGLVDAAEY 1023  
 DB SRPRFRELITEFSMAADPQRYVLIQDDEHRLPSPTDSNRYRALMOEDMDVDVDADEY 1016  
 QY LVPQGGFCPPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEGAS 1083  
 DB LVPQGGFCPPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEGAS 1042  
 QY LIPQGGFCPPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEGAS 1042  
 DB LIPQGGFCPPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEGAS 1042  
 QY DVFPGDGLGMAAGLQSLPETHDPSLQRYSEDPVLPSET--DGYVAAPLTCSPQPEYVN 1141  
 DB DVFPGDGLGMAAGLQSLPETHDPSLQRYSEDPVLPSET--DGYVAAPLTCSPQPEYVN 1094  
 QY N--NSTVACIDRNGLOSCPIKEDSFQRYSSDPGALTEDSIDDTFL-----VPEYIN 1094  
 DB N--NSTVACIDRNGLOSCPIKEDSFQRYSSDPGALTEDSIDDTFL-----VPEYIN 1094

```

1142 QPDVRFPPSPREGELPAAPBAGATLERAKTLPSPKGVVYKDVAFAGAVENPEYL-TPQ 1200
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1095 Q-SVPRPRPGSVQNVNHNQNLN-APSRDPHYD--PHSTVGVNPEYLNTVQ 1143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1201 GGAAPQHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPSPFFKGTPTAE 1244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1144 -----PTCVNSTFSDSPAHMAQKSHQISLNDPVDQODFFPKAKKPKNGI-FKGS-TAE 1197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1245 NPEYL 1249
      : : : : :
1194 NAEYL 1198

```

STANDARD:	PRT;	1210 AA.
EGFR MOUSE		
MOUSE		
5		

001279; 33, Created)  
01-FEB-1996 (Rel. 33, Last sequence update)  
01-FEB-1996 (Rel. 41, Last annotation update)  
15-JUN-2002 (Rel. 41, Last annotation precursor (EC 2.7.1.112).  
Epidermal growth factor receptor  
ECFR.  
Mus musculus (Mouse)  
Mus musculus; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia;  
NCBI TaxID=10090;

FROM N.A.

SEQUENCE: TISSUE=LIVER;  
STRAIN=BALB/c; PubMed=1408137;  
Accession: 33026370; Givol D.;  
with factor receptor 2

Avivi A., Skorecki K., Yayon A. MEDLINE=3508000  
"Promoter region of the murine fibroblast growth factor-1 (bek/KGFR) gene." J Biol Chem 267:7195-7197, 1992.

Oncogene 1992; 10: 111-120

SEQUENCE FROM N... TISSUE=LIVER  
STRAIN=BALB/c, and CD-1; PubMed=7678348;  
New S.K.;

MEDLINE=93126380; Andrews G.K., Das S.K., paria B.C., "The effect of the epidermal growth factor receptor on the timing of delayed implantation."

"Expression of  $\alpha$ -crystallin in mouse blastocysts during development." *Dev. Biol.* 161: 1-11. Acad. Sci. U.S.A. 90:55-59 (1993)

Proc. Natl. Acad. Sci. U.S.A. 73: 1000-1004, 1976.

SEQUENCE FROM N.5.  
STRAIN=BALB/c; TISSUE=Liver;

Hibbs M.L.;  
submitted (APR-1994) to the EMBL/GenBank/DBS

RL  
Submitted  
(4) FROM N.A.

RP TISSUE=Liver;  
SEQUENCE FROM:  
STRAIN=B6/C3J; T H Copeland N.G., Earp H.S.,  
19986. PubMed=8125255;

MEDLINE=94170980, Phillips H.K., Qiu L., et al. Twelve N.C., Phillips H.K., Qiu L., et al. Even a point mutation in the EGF

Lee D.C.: 'Jenkins N.A.: phenotypic mouse waved-2 phenotypic mouse'

RT "line model"  
RT receptor tyrosine kinase.  
8.399-413 (1994)

Gene	Dev.	From N.A.
RL	[5]	1-714
RN		

SEQUENCE OF 1-72  
RP TISSUE=Brain; PubMed=2030916;  
Givol D., Morse B.;

RC MEDLINE=91232866; Fulrich A., Schesbinger-  
RX MEDLINE=91232866; Fulrich A., Schesbinger-  
Medi A., Lax I., ... sequences as a guide to study the ...

### RT Comparison of EGF receptor

RT	binding
RL	Oncogene 6:673-676 (1991).

[6] SEQUENCE OF 969-1117 FROM N.A.  
RN  
PD

STRAIN=C3H; Serrero G.; to the EMBL/GenBank/DBJ databases.  
 risinger D.P.; 1990) and the BIOLOGICAL SIGNAL OF EGF, CP30 ANT

Submitted (JUN-1992) to **EGF RECEPTOR MEDIATES IN**  
**FUNCTION: THE EGF RECEPTOR MEDIATES, HEPARIN-BINDING EGF,**  
**AMPHIREGULIN, HEPARIN-BINDING EGF,**

-10 FUNCTIONALITY OF THE  
CC AND OF TGF- $\alpha$ ,  
AND ALSO OF GROWTH FACTOR (BY SIMILARITY)  
CC AND VIRUS GROWTH = ADP + protein tyrosine

CC tyrosine phosphate. Type I membrane protein.  
CC SUBCELLULAR LOCATION: Type I receptor leads to  
-1- SUBCELLULAR LOCATION: Binding of EGF to the receptor complex,  
CC -1- MISCELLANEOUS: Internalization of the EGF-receptor complex,  
CC dimerization, internalization of the EGF-receptor complex,  
CC induction of the tyrosine kinase activity, stimulation of cell DNA  
CC synthesis, and cell proliferation.  
CC INDUCTION OF THE TYROSINE KINASE ACTIVITY:  
-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at  
CC the European Bioinformatics Institute. There are no restrictions on its use  
CC by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC [license@isb-sib.ch](mailto:license@isb-sib.ch)) .  
-----

CC	or send an email to lrc@ncbi.nlm.nih.gov
CC	-----
CC	EMBL; X78987; CAA55587.1; -
DR	EMBL; U03425; AAA17899.1; -
DR	EMBL; X59696; CAA42219.1; -
DR	EMBL; L06864; AAA53029.1; -
DR	EMBL; Z12608; CAA78249.1; -
DR	HSSP; P11362; IEGF.
DR	MGD; MG1.95294; Egfr. EGFR_L domain.
DR	InterPro; IPR000494; EGFR_kinase.
DR	InterPro; IPR000719; Euk_kinase.
DR	InterPro; IPR002174; Furin-like.
DR	InterPro; IPR001245; Tyr_kinase.
DR	Pfam; PF00069; pkinese; 1.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PR01030; Recep_L_domain; 2.
DR	Pfam; PD000001; Euk_pkinase; 1.
DR	ProDom; PD00261; FU; 3.
DR	SMART; SMO0259; TyrcK; 1. KINASE ATP; 1.
DR	SMART; SMO0107; PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR	PROSITE; PS50011; Receptor; Signal; Transferase.
DR	Transmembrane; Glycoprotein; Repeat.
KW	transmembrane-protein kinase; ATP-binding; POTENTIAL.
KW	Tyrosine-kinase; 24
FT	SIGNAL 1 1210
FT	CHAIN 25 647
FT	DOMAIN 25 670
FT	TRANSMEM 648 1210
FT	DOMAIN 671 1210
FT	REPEAT 75 300
FT	REPEAT 390 600
FT	DOMAIN 1028 1071
FT	DOMAIN 714 728
FT	NP_BIND 720 747
FT	BINDING 747 839
FT	ACT_SITE 839 199
FT	DISULFID 190 207
FT	DISULFID 194 223
FT	DISULFID 215 231
FT	DISULFID 219 240
FT	DISULFID 232 248
FT	DISULFID 236 260
FT	DISULFID 251 261
FT	DISULFID 264 307
FT	DISULFID 295 326
FT	DISULFID 311 333
FT	DISULFID 329 333
FT	DISULFID 506 523
FT	DISULFID 510 535
FT	DISULFID 526 555
FT	DISULFID 539 555
FT	DISULFID 558 571
FT	DISULFID 562 579
FT	DISULFID 582 591
FT	DISULFID 595 617
FT	DISULFID 620 628
FT	DISULFID 620 636
FT	MOD_RES 684 680
FT	MOD_RES 1092 1092
FT	MOD_RES 1092 1092
FT	PHOSPHORYLATION (BY PKC) (BY SIMILARITY)
FT	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT	MOD_RES	1110	1110	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	MOD_RES	1172	1172	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	MOD_RES	1197	1197	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT				(BY SIMILARITY).	
FT	CARBOHYD	128	128	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	175	175	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	413	413	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	444	444	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	568	568	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	603	603	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	19	19	C -> S (IN REF. 2).	
FT	CONFLICT	539	539	C -> M (IN REF. 5).	
FT	CONFLICT	991	991	L -> F (IN REF. 4).	
FT	CONFLICT	1116	1117	HP -> DR (IN REF. 6).	
QO	SEQUENCE	1210	134853	MM; 690E20D46DFD2D3F5	CRC64;

Query Match	45.0%	Score 3053	DB 1	Length 1210
Best Local Similarity	49.0%	Pred No. 6.5e-158		
Matches 623	Conservative 159	Mismatches 370	Indels 110	Gaps 23

QY	1	LLALLLRPGA--	STOVGTGDMKRLRLASPEHLLDMRLHLYOGGQVVOGNNELTYLPTN	68
Db	14	LLTLLCAAGGLAEKKVCQGTSNRLTOLGTETDHLSLQRMKNCEVVLGNLEITYVORN		73
QY	69	ASLSFLADIQEVQGVLLIAHNOVROVPLORIRIVGTOLFEQNVATLAVLNDQPLNNTTP		128
Db	74	YDLSFLKTIQEVAGVLLIANTVEIRPLENQITIGNALIKENTYATLALSN-----		124
QY	129	VTGASPGGLRELQRLSLTEILKGGVLLQANPOLCYODTILWKDI---FHKNQALATLI		184
Db	125	-YGNTRGLRELPMRNLQELIGARFSNNPILCMNDTIQWRDIIQVNFMSMSMDL---		180
QY	185	DTNRSBACHPCSPMKSGRCMGESEEDQSLRTYCAAGCA-RCKGRPLQYIKANSKITG		243
Db	181	-QSPHSCTPKDPCSPNGSCWGGEGENCQKLTIIQAQCSHRCKGRBPSDCHQCAAG		239
QY	244	ITELKHSDDLCLAFHNHSGICELHCPALVTNTDFEEMPNEGRYTGASCVTACPNY		303
Db	240	CTGRRESDDLVCQKFOBATCKOTCPRLMLNPTTYQMDVNPBGKTSIGATCYKKCPNY		299
QY	304	LSTDVGSCTVCPRLHNOEVTAEADGTQRCSEKSKPCARCYGLMEHLEEVANISANTQE		363
Db	300	VVTDHGSQVRAQGRDYEV-EDGIRKCKKCDGPRCKYONGIGIEFDTILSNANTIKH		358
QY	364	PAGCKKTFGSLAFLEPESFDGPASNTAPLQREBOLVFETLEITGYLYISAMPDLS		423
Db	359	FKYCTALISGDHLIPLVAFKGSFTRTPDLRELEITLKYKEITGFLIIQAMPDWTDLH		418
QY	424	VFOULQVIRGILHNGASVLLQGLGSIWLSRLRLRELQSGUALIHNTHLCFVNTVWD		483
Db	419	AFENLEIRIGTKOHGQPSLAVVGNTISLGRISKEISDDVILISGRNLQYANTIMWK		478
QY	484	QLFENPQALHTANRPEDECVCGEGLACHOLCAHGCWGPPTQCVNCSQFLRGQCEVEE		543
Db	479	KLFTQPNQKTIIMNBRKEDCKAVNHYVNPILCSBGSCMGPRDQVCSQVNSRGREVEK		538
QY	544	CRVLQGLPREVYNAHCLPCHPEQOPONGSVTCFGEADQCVACAHYKDPFCVAPRCBG		603
Db	539	CNILEGEPREVEVSECIQCHPECLPQAMNITITGRGPDNCIQCAHYIDGPHCVTCDAG		598
QY	604	VKPDLSIMPIIMKPFDEBGAQCPCEIINCNSCVDDDKCQRAEQRASPLTISVAVVGLL		663
Db	599	IMGENNTLL-WKIVADANNVCHLCANCTVGGAGPGLQCEVMPGSPKLSIATGVGLL		657
QY	664	VVVLGVAVFGI-LIRROOKIRKTYMRRLQTELEVEPLTPSGGAMPNQOMKILKETEARK		722
Db	658	FIIV-VALGICLPMRRKHIVAKTLIRRLQRELEVEPLTPSGEARNQAHRLILKETEKK		716
QY	723	VKVLGSAFCTVYKGIIPDGENVKIPVALIKVLEKNTSPKANKTEILDEAVYAGVGSYV		782

[illegible]

## RESULT

ID	ERB4_HUMAN	STANDARD;	PRT;	1308 AA.
----	------------	-----------	------	----------

DT	15-DEC-1998	37	1205 000
21	15-DEC-1998	(REL. 37,	CREATED/

Receptor protein-tyrosine kinase

GN ERBB4 OR HER4.

Mammalia: Eutheria: Primates: Platyrrhini: Cebidae: Cebus: Cebus

RN [1]

ISSUE=Breast carcinoma;

BA For T. Neubauer M G Shoyab

RT epidermal growth factor receptor

23 [2] 23

RX MEDLINE=97476287: PubMed=93342

RA Klagsbrun M.;

RT  
VI  
CROSS REPRODUCTION AND DIRECT  
REPRODUCTION " "

[illegible]



```

Db      536 CNLYDGEFREFECSICVECDPCECKMEDGLTCHGPGPNDCTKSHFKDGNCEKCPD 595
Qy      603 GVKPDLSPYMPKPEDEGACQPCPINCTHSCVDLDDKGC-----PAEQRASPL 651
Db      596 GLOQANF--IFKXADPRECHPCHPNCTOGCNPSTSDCIYPTWGTSTLPQAR--TPL 652
Qy      652 TSIVSAVV-GILLVVLGVFGLIKRRQOKIRKTYMRLLQETLVEPLTPSGAMPNQA 710
Db      653 --IAAGVIGGLFLLIVIGLTFEAVVYRRRSIK-KKALRRFL-ETELVEPLTPSGAPNQA 708
Qy      711 QMRLLKTELKRVKLVLSGAGCTYKKGIMIPDGENVKIPVAIKVLRENTSKANKEILDE 770
Db      709 QLRLLKTELKRVKLVLSGAGCTYKKGIMIPDGENVKIPVAIKVLRENTSKANKEILDE 768
Qy      771 AYWAGVSPYVSRLLGLCTSTVOLVQLMPYGLLDHVENRRLSOLLNMCQIJA 830
Db      769 ALIAMSADHRLVLLGVCLSPITQVQLMPHGLLEYVHNKDNISOLLNMCQIJA 828
Qy      831 KMSYLEDVRLVHRDLAARNLVKSPNHVKTITDGLARLDIDETEHADGKVPKXMA 890
Db      829 KGMVYLEERLVRHDLAARNLVKSPNHVKTITDGLARLDIDETEHADGKVPKXMA 888
Qy      891 LESTLRBFTHOSVMSYGVTEMLMTGAKPYDGIAPARELPDLEKGERLPQPICTID 950
Db      889 LECIHYRKFTHOSVMSYGVTEMLMTGAKPYDGIAPARELPDLEKGERLPQPICTID 948
Qy      951 VYIMIVKCMWIDSECRPFRELVESEFMRMADPQRFVVIQONED-LGPASPLDSTFYRSL 1009
Db      949 VYIMIVKCMWIDSECRPFRELVESEFMRMADPQRFVVIQONED-LGPASPLDSTFYRSL 1008
Qy      1010 EDDMDGLVDAAEYLVPOQGFCCDPAPAGAGMHHRRSSSTSGGDLTLGLEPSEE 1069
Db      1009 DEEDLEMDMAEEYLVLP-QAFNIIPP-----IYTSRARIIDSNRS-----EIGHSPPAY 1056
Qy      1070 APRS-----PLAP-SEGSGSVDFDGLDMGAKGLQS 1100
Db      1057 TPNSGNPFYVRDGFPAEEGVSVPRAPFTSTIPEPVAQAGTAELFDSGCCNGLTRKXVA 1116
Qy      1101 LPTHDPSPLQRYSEDPVPLPS-----ETDGVAPLTCSPQPEYVQDPVRQPPSPR 1153
Db      1117 PHVEDSSTQRYSDPTVYFAPERSPRGELDEGVTMPRDKPKQEIYLPV----- 1167
Qy      1154 EGPLPAARPGATLERAKTSLSPKNGGVKDVFAFGAVENPEYLTPOGGAAPQPHPPA- 1212
Db      1168 ENPVSRR-----KNGDLQ-----ALNDENYHNASNG-----PRYAE 1199
Qy      1213 -----FSPAFDNLYMDQPPERGA--PPSTF 1237
Db      1200 DEYVNEPLYLNTFANTLGKAEYLNKNTLSMPEKAKKAFDNDYWNHSLPPRSTLQHPDYL 1259
Qy      1238 KGTPT-----AENPEYL 1249
Db      1260 QEYSTKFFYKONGRIPIVAENPEYL 1285

```

```

RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      (2)
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=91222560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system."
RL      Neuron 6:691-704(1991).
RN      (3)
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frohmet P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulins and their putative receptors, ErbB2 and
RT      ErbB3, is induced during Wallerian degeneration."
RL      J. Neurosci. 17:1642-1659(1997)
CC      -I- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NR-
CC      2, NR-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -I- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -I- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -I- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as there is no content in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF041838; AAD08899.1; -.
DR      EMBL; U52531; AAC53051.1; -.
DR      HSSP; P11362; IFGK.
DR      InterPro; IPR000494; EGFR_L domain.
DR      InterPro; IPR000719; Euk_Pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_Pkinase.
DR      InterPro; IPR004019; YLP motif.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF00069; Pkinase; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      Pfam; PF02757; YLP; 2.
DR      PRINTS; PR00109; TYRKINASE.
DR      PRODOM; PD000001; Euk_Pkinase; 1.
DR      SMART; SM00261; FU; 4.
DR      SMART; SM00219; TYKIC; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      PROSITE; PS0109; PROTEIN_KINASE_TYR; 1.
KW      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal.
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT      CHAIN 1 25
FT      DOMAIN 26 1308
FT      TRANSMEM 652 675
FT      DOMAIN 676 1308

```



Query Match 42.6%; Score 2890; DB 1; Length 1308;  
Best Local Similarity 44.8%; Pred. No. 4,7e-149;  
Matches 603; Conservative 189; Mismatches 395; Indels 160; Gaps 28;  
SQ SEQUENCE 1308 AA; 146957 MW; D944B096A08B41 CRC64;  
FT DOMAIN 186 334 CYS-RICH.  
FT DOMAIN 496 633 CYS-RICH.  
FT DOMAIN 718 985 PROTEIN KINASE  
FT NP BIND 724 732 ATP (BY SIMILARITY).  
FT BINDING 751 751 ATP (BY SIMILARITY).  
FT ACT\_SITE 843 843 BY SIMILARITY.  
FT DISULFID 189 197 BY SIMILARITY.  
FT DISULFID 193 205 BY SIMILARITY.  
FT DISULFID 213 221 BY SIMILARITY.  
FT DISULFID 217 229 BY SIMILARITY.  
FT DISULFID 230 238 BY SIMILARITY.  
FT DISULFID 234 246 BY SIMILARITY.  
FT DISULFID 249 258 BY SIMILARITY.  
FT DISULFID 262 289 BY SIMILARITY.  
FT DISULFID 293 304 BY SIMILARITY.  
FT DISULFID 308 323 BY SIMILARITY.  
FT DISULFID 326 330 BY SIMILARITY.  
FT DISULFID 503 512 BY SIMILARITY.  
FT DISULFID 507 520 BY SIMILARITY.  
FT DISULFID 523 532 BY SIMILARITY.  
FT DISULFID 536 552 BY SIMILARITY.  
FT DISULFID 555 569 BY SIMILARITY.  
FT DISULFID 559 577 BY SIMILARITY.  
FT DISULFID 580 589 BY SIMILARITY.  
FT DISULFID 593 614 BY SIMILARITY.  
FT DISULFID 617 625 BY SIMILARITY.  
FT DISULFID 621 633 BY SIMILARITY.  
FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-).  
FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-).  
FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-).  
FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-).  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 1062 1062 S -> N (IN REF. 3).  
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).  
Query Match 42.6%; Score 2890; DB 1; Length 1308;  
Best Local Similarity 44.8%; Pred. No. 4,7e-149;  
Matches 603; Conservative 189; Mismatches 395; Indels 160; Gaps 28;

QY 357 TSANIQFAGCKKIFGSLAFLPESFQDPASTAPLQEQLOVFETLEITGYLISAMP 416  
DB 349 DSSNDKINCKIKINGNIFLVTGKHGDPYNAIDIDBEKLVFTVETIEITGFLNIQTMR 408  
QY 417 DSLPDLVFNQLOVIRGRIHLNGAVSLTLOGISWGLRLSRLREGSGSLAIHNHTLCE 476  
DB 409 PNMDFSVFSLVLTIGRVLVSGSLILTIKQGGITSLQFOSIKEISAGNIYITDNSNLCT 468  
QY 477 VHTPMDLFRNPHQALHTANRPDECEVGLACHOCARGCHGPRPTQVNCQSOFRL 536  
DB 469 YHTIMWTLFTVQRIYIRDNRRACNCTAGCMVGNHLCSDNGCWPBPDOCLSCRFSR 528  
QY 537 GOECVEECRVYQGLREYVNAHCLPCHRECP -ONGSVTCGCPADOCVACAHKXDPF 595  
DB 529 GKICIESGNLVDGEREFENGSIYECESQCEKMEGDLTCHGPRPDNCTKSHRKDGN 588  
QY 596 CVARCPGVKPDLSYMPIWKEPDEBGACQPCPINTHSCVLDLDDKGC-----PA 644  
DB 589 CVEKCPVDLQANSE--IFKYADQRECHPCHPNCTGCGNGPSTSHDCIYYPMTGSHTLPO 646  
QY 645 EORASPLTISYSAV--GILLVVVLCGVFGLIKRQOKIKRTYMRLLQETELVETPS 703  
DB 647 HAR--TPL--IAAGVIGGLFVLVIMALTFAVYVRRSIR-KKRALRFL-ETELVETPS 701  
QY 704 GAMPQAOIRLIKETELRKVKVLSGAGFQYVKGIMIPDGENVKIPVAIKVLRNTSPKA 763  
DB 702 GTAPRQAOIRLIKETELRKVKVLSGAGFQYVKGIMIPDGENVKIPVAIKVLRNTSPKA 761  
QY 764 NKEILDEAYVAVGSPYVSRLLGICLTSTVOLVTLMPYGGCLLDHVENRGLCSODLL 823  
DB 762 NVEFDEALIMASVDHPHVLVGLVCLSPITOLVQLMPHGCLLEYVHEHKNIGSOLL 821  
QY 824 NMCQIAGMSYLEVRVLRHDLAARNLVYSPNNVKTITDGLALLDIDEFYAADGK 883  
DB 822 NMCQIAGMSYLEVRVLRHDLAARNLVYSPNNVKTITDGLALLDIDEFYAADGK 881  
QY 884 VPIKMALESILRRFTQSDVWSYGVTVMELTGAKPYDGI PRAEIPDLLEKGERLPQ 943  
DB 882 MPKIMALECHYKRFTHQSDVWSYGVTVMELTGAKPYDGI PRAEIPDLLEKGERLPQ 941  
QY 944 PPICTIDVYIMVCKMIDSECRPFRELVSFESMABDPORFVYIIONED-LGPASPLDS 1002  
DB 942 PPICTIDVYIMVCKMIDSECRPFRELVSFESMABDPORFVYIIONED-LGPASPLDS 1001  
QY 1003 TFRSLEDDMDGLVDAEYLVLPQGFPCDP-----ARCA 1039  
DB 1002 KFFQVLDLEEDLEMDMAEYLVLP-QAFNIPPIYTSRTIRDSNRSEIGHSPPAYTPMS 1060  
QY 1040 GAMVHRRSSSTRSGGDLTLGLEPSEEBAPRSPLAPSEGADVPODLCMGAKAGLQ 1099  
DB 1061 GSGPVLQDGGFATQGG---MEMPTATSTIRBAVLA--QATAMFDDSCNGTLRKPV 1115  
QY 1100 SLPTHDPSPLORYSEDPVPLPS-----ETDGVAVRLTSCPOREYVQNPDRPQPS 1152  
DB 1116 VPHVQEDSSTQRYSDPVPFAPERNRPAELDEBEGVTMHRPKQDEYLNPAVE----- 1167  
QY 1153 REGPLAPRAGATIELRAKTLSPCKNGVYKQVAFAGCAVENEPEYLTTPGGAAPQHPRA 1212  
DB 1168 -ENPVSRR-----KMGDLQ-----ALDNPETHSASG-----PPA 1198  
QY 1213 -----FSPAEDNLVYWDOPPERGA--PPST 1236  
DB 1199 EDEYVNEPLVANTFTNALGNAEYMGNSLSPBEKAKKAFNDPVDVYNNHSLPFRSTLQHPDY 1258  
QY 1237 FKGTPT-----AENPEYL 1249  
DB 1259 LOEYSTKYFYKONGRIRPIVAENPEYL 1285  
RESULT 8  
XMRK\_XIPWA  
ID XMRK\_XIPWA STANDARD; PRT; 1167 AA.  
AC P13386;



DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).  
 OS XMRK OR TU.  
 OS Xiphophorus maculatus (Southern platyfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 NCBI\_TaxId=8083;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90015140; PubMed=2797166;  
 RA Mittlebrot J., Adam D., Malitschek B., Mauelel W., Raulf F.,  
 RA Telling A., Robertson S.M., Scharlt M.;  
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-  
 RT inducing Tu locus in Xiphophorus.";  
 RL Nature 341:415-421(1998).  
 RN [2]  
 RP REVISION TO 515.  
 RA Scharlt M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; X16891; CA34770.2; -  
 DR PIR; S06142; S06142.  
 DR HSSP; P11362; 1FGK.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recept\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 5.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR SMART; SM00219; TY-TKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.  
 FT CHAIN 1 25  
 FT SIGNAL 1 25  
 FT 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE  
 FT KINASE.  
 FT DOMAIN 26 642  
 FT TRANSMEM 643 665  
 FT POTENTIAL.  
 FT DOMAIN 666 1167  
 FT CYTOPLASMIC (POTENTIAL).  
 FT NP BIND 710 977  
 FT PROTEIN KINASE.  
 FT BINDING 716 724  
 FT ATP (BY SIMILARITY).  
 FT ACT\_SITE 743 743  
 FT BY SIMILARITY.  
 FT DISULFID 195 204  
 FT BY SIMILARITY.  
 FT DISULFID 220 228  
 FT BY SIMILARITY.  
 FT DISULFID 224 236  
 FT BY SIMILARITY.

FT DISULFID 237 245 BY SIMILARITY.  
 FT DISULFID 241 253 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT DISULFID 269 296 BY SIMILARITY.  
 FT DISULFID 300 311 BY SIMILARITY.  
 FT DISULFID 315 330 BY SIMILARITY.  
 FT DISULFID 333 337 BY SIMILARITY.  
 FT DISULFID 504 513 BY SIMILARITY.  
 FT DISULFID 508 521 BY SIMILARITY.  
 FT DISULFID 524 533 BY SIMILARITY.  
 FT DISULFID 537 553 BY SIMILARITY.  
 FT DISULFID 556 569 BY SIMILARITY.  
 FT DISULFID 560 577 BY SIMILARITY.  
 FT DISULFID 593 615 BY SIMILARITY.  
 FT DISULFID 618 626 BY SIMILARITY.  
 FT DISULFID 622 634 BY SIMILARITY.  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;  
 Query Match 38.7%; Score 2625.5; DB 1; Length 1167;  
 Best Local Similarity 44.8%; Pred. No. 8.5e-135;  
 Matches 568; Conservative 162; Mismatches 400; Indels 137; Gaps 27;  
 Oy 4 AALCRMGILLALLPFGAAS- - - - - QVCTGDMKRLRSPETHLDMRLHLCGGCOYOGN 59  
 8 AALIQ--LLVLVLSIRCCSTDPDRKVCQGSINQTM--LDNHLVKKMKMYSGCCNVLVEN 62  
 Oy 60 LELTYLPTNASTFLDIOIEVOGYVLLAHNOVROVPLQRLRIVRGTOLEFDNVALAVLDN 119  
 63 LEITYQENDLSTLQSGVGVLLAMEVSTPLVNLRLINGONLEGNFTLLVWSN 122  
 Oy 120 GDPPLNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNL 179  
 123 YQK-NPSSP--DYQVGLKQLOLSNLTIELISGGYKVSINPLLCVETIMMDIVDKTSNP 179  
 Oy 180 ALTLIDNRSRACHPCSPMCKGSRCKWSSSDCSQSLRTVAGGC-ARCKRPLQYITAN 238  
 180 TMLNLIPIAFERQCKCHGCVNGSCMAPRGHCQKFTKLCAEQCNRRCRGPKRPDDCNE 239  
 Oy 239 SKFIGITELKHSDDLACLFHNSGICELHCPALVTYNTDFESMNPREGRTYFGASCVTA 298  
 240 HCAAGCTGPRATDCLACRDFNDSTCTCDTPRPRIYDVSQVVDNPNIKITTFGAACVKE 299  
 Oy 299 CPYNYLSTDVSGCTLVCPRLNQEVTAEVGTORECKSKPCARVCYGLGEMHLEVRVATVS 358  
 300 CPSSNYVYTE-GACVRSACAGLEVD-ENGKRSCKPCDGVCPKVDGIGISLSTIAVNS 357  
 Oy 359 ANIOEPAGCKKIGSLAFLPESFPDGPASNTAPLPQPOLQVFETLEITGLYISAMPDS 418  
 358 TNIRSFNCKTKINGDILLNPNSEFGDHPYIKGTMDPRLNLTIVKTEITGLVIVMMPEN 417  
 Oy 419 LPDLSPVQNTQVIRGRILHNGAVS-LTLQGLGISMLGLRSRLSGSLALHNHTHJCFLV 477  
 418 MTSLSVQNTLEIRGRITTFRGRFSFVVVYRHLQMLRLSKVSAAGVILAKTLQRLRYA 477  
 Oy 478 HTVPMQDLPFNPHQALHTNRPBDECVGGLACHQLCARGHCKGPPPTQCVNCSQFLRG 537  
 478 NITNMRILFRSEDSIEYDART-----ENQTCNNESBGSCGPPPTMCSCLHVDRG 530  
 Oy 538 QECVEGRVYQGLPREVYNARHCLPCHREGCPONGSTYCGPREDDQCAAHYDDPPFCV 597  
 531 GRCVASCNLQGRERQVDGRVCQHQECCLVGTDTSLTYGPREPANSKSAHFQDGQCI 590  
 Oy 598 ARCPGKVPDLSTYPIWKFPDEGACORPCPINCTNSCVDDDBDKGCPRAEGRASPLTISVSA 657

```

Db      591 PRCHGILGDDDLT-IMKXADMGCCQCHONCTGCGCGPGLSGGRD-IYSHSLAAGL 648
      658 VVGLLVVGLVGVGILIKRROOKTKYTMRLLOETELVEPLTSGAMPNOQRIIKE 717
      649 VSGLLITVIVALLVLLRRRRRIK-RKRTICLLLOEKELVEPLTSGQAPNOAFRIIKE 707
      718 TELRKVVLGSGAGCTYVKGIWIPGGENVKIPVATKVLRENSPANELLIDEAVVMGV 777
      708 TEFKDRVLGSGAGCTYVKGLMNPGENIRIPVATKVLREATSPVNOEVLDEAVVMASV 767
      778 GSPVYSRLIGICTSTVOLVTLQMLPYGCLLDHVRNRRGLSGODLLAMCMQIAKMSYLE 837
      768 DHPHVCRLIGICTSAVOLVTLQMLPYGCLLDVVRHQRIRICQWMLNCVOIAKAMNLE 827
      838 DVLVHRLAARNVLYKSPNHYKITDGLARLLIDETRYHADGGRVPIKMALESIIIR 897
      828 ERHLVHRDLAARNVLLKNPNHYKITDFGLSKLLTADKEKEYQADGGKVDIKMALESIIQW 887
      898 RFTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIDLLEKGRLLPOPICTIDVYMIWK 957
      888 TTHOSDVWSYGVTVWELMTFGSKPYDGIIPAKELASVLENGERLPOPICTILEVMIILK 947
      958 CWMIDSECRPFRELVSFESRWARDPQRFVIVONEDLGPASPJLSTFYRSLLEDDMGDL 1017
      948 CWMIDPSRPRFRELVSFESQWARDPSRYLVIG--NLPSLSDRLLFSRLSSD--DV 1002
      1018 VDAEYLVLPQGFPCFPDPAFAGAGVHHRHSSSTRSGGDLTGLPSEEEAPSPILAP 1077
      1003 VDAEYLLPYKRI-----NRQSS-----EPCIP 1025
      1078 SEGAGSDVFDGDLGMAKAGLSLPTHDPSPLOQRYSEPTV-PLPSETGVYAPLTCSPQ 1136
      1026 PTGH-----PVRENSITLRISPTONALKDLDG----- 1056
      1137 PEYVNPDPVRPOP-----PSPRE-----GLP-AAPAGATLERAKTLSPGKGVYKD 1183
      1057 -EYVNPQSESTSRSLDIYNPYEDLTDQMGVSLSGQEAENFSPREVLINQNSL--- 1112
      1184 VFAFGAVENPEYLTPOGGAAPQHPHPAFSPAFNLYYWDODPPERGAIPSTFGTPTA 1243
      1113 PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTALTGNGMFLPAA 1150
      1244 ENPEYIG 1250
      1151 ENLEYLG 1157

```

## RESULT 9

ERB3\_HUMAN STANDARD: PRT: 1342 AA.

```

ID_ERB3_HUMAN
AC P21850.
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Klaus M.H., Issing W., Miki T., Pinescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

```

```

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTK.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC - ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC - TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC - LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC - DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M29366; AAA5790.1; -.
DR EMBL, M34309; AAA55979.1; -.
DR EMBL, S61953; AAB26935.1; -.
DR PIR, A36223; A36223.
DR HSSP, P11362; 1FGK.
DR GeneW, HGNC:3431; ERBB3.
DR MIM, 190151; -.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam, PF00069; Pkinase; 1.
DR Pfam, PF00757; Furin-like; 1.
DR Pfam, PF01030; Recep_L domain; 2.
DR ProDom, PD000001; Euk_kinase; 1.
DR SMART, SM00261; Fu_3.
DR SMART, SM00219; TyKc; 1.
DR PROSITE, PS00107; PROTEIN_KINASE_ATP; FALSE NEG.
DR PROSITE, PS00109; PROTEIN_KINASE_TYR; FALSE NEG.
DR PROSITE, PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1..19
FT CHAIN 20..1342
FT DOMAIN 643..664
FT TRANSMEM 644..664
FT DOMAIN 665..1342
FT DOMAIN 709..966
FT NP_BIND 715..723
FT BINDING 742..742
FT ACT_SITE 834..834
FT DISULFID 186..194
FT DISULFID 190..202

```

```
FT DISULFD 210 218 BY SIMILARITY.
FT DISULFD 214 226 BY SIMILARITY.
FT DISULFD 227 235 BY SIMILARITY.
FT DISULFD 231 243 BY SIMILARITY.
FT DISULFD 246 255 BY SIMILARITY.
FT DISULFD 259 286 BY SIMILARITY.
FT DISULFD 290 301 BY SIMILARITY.
FT DISULFD 305 320 BY SIMILARITY.
FT DISULFD 323 327 BY SIMILARITY.
FT DISULFD 500 509 BY SIMILARITY.
FT DISULFD 504 517 BY SIMILARITY.
FT DISULFD 520 529 BY SIMILARITY.
FT DISULFD 533 549 BY SIMILARITY.
FT DISULFD 552 565 BY SIMILARITY.
FT DISULFD 556 573 BY SIMILARITY.
FT DISULFD 576 585 BY SIMILARITY.
FT DISULFD 589 610 BY SIMILARITY.
FT DISULFD 613 621 BY SIMILARITY.
FT DISULFD 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 ELISGVYIEKDKLCHMDITDMRDIVDRDEIYKONGR
SC -> GQPMVPSGLTQPMADWYLDLDDPDLTLTSSSK
VPTLAIV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E766CA374BD CRC64;
```

Query Match 34.7%; Score 2354.5; DB 1; Length 1342;  
Best Local Similarity 40.1%; Pred. No. 4.6e-120;  
Matches 526; Conservative 190; Mismatches 466; Indels 129; Gaps 32;

```
10 GLLTALLPRGAA--STQVCTGTDMLRLPAPERTHLDMLRYGCGGVQGNLEIYPT 67
11 GLTSLARGSEVNSQAVCPGLTGLSTGDAENQYOTLYKLYERCEVVMGNLEIYVTGH 70
68 NASISFLDIQEVGYVLIANOVQVPLQRLRIYRGTLFEDNVALLAVLNDPILNNT 127
71 NADISFLQWIREVGYVLIANNEFTLPLPILRYVRGTQVYDGFALITVM-----LNT 125
128 PVTGASPGGLRELIQLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNOALATLIDTN 187
126 ----NSSHALRQLRLTOLTEILSGVYIEKNDKLCMDITDMRDIVDRD---AEIVKXD 178
188 RSPACHSPCKSRCKSGESSEDCOSLTRTVACGGC-ARCKGRLPYOTIKANSKFTIGITE 246
179 NGRSCPCHEVCCKG-KCMGPSEDCQTLTKTICAPQCCHGFCGPNPNCCHDEACAGCGSG 237
247 LKISDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYFGASCVTACPNYLTST 306
238 PQRDIDCACHHFNDSGACVPRCPPLVYNTKTLFQLEPRPHKTYQYGGVCAVSCPHNFV-V 296
307 DVGSGTIVCPLNHNEVTAEDGTORCEKSKPCARVCGYLGMEHLREVAVTSANIQEFAG 366
297 DQTSVCAVACPDDKMEVD-KNGLKMCCEPCGGLCPKACEGTGSG--SRFTVDSNIDGFVN 353
367 CKKIFGSLAPFESFDDPASNTAPLOPEQOVFETLEETIGYLYISAMPISLDLSYFQ 426
354 CTILGLNLDPLITGLNDPWHKIPALDPEKLVNRTVARELITGYLNIOSWPHMENFSVFS 413
427 NLQVIRGRIILHNGAYS-LTLQGLGISWLGRLSRLRELSGLALIHNNTHLCPVHVPMDOL 485
414 NLTTIGRSLYNNGFSLIMKMLNVTSLSGFRSLKEISAGRIYISANQOLCTHHSMLNMTKV 473
```

```
QY 486 FENPQALHTA-NREDECVGEGLACHQLCARGHGMCGPTQVNCVSOLFKEQECVEEC 544
DB 474 LRGPTEERLDIGNRPRRCVCAVEKVCDBPLSSGGCGMGPGQGLSCSRNYSKGVCTHIC 533
QY 545 RVLQGLPREYVNAHRLCPHPECOPQNSVTCFGBADQVACAHYKDPFVCARCPSGV 604
DB 534 NFLNCEPREFAHEACFCSCHEPCQMEGTATCGSGSDTCAOAHFRDQPHCVSCPHGV 593
QY 605 KPDLSYMPIMKPEDEGACORPCINCTHSQVLDLDDKGCABQRA-----SPLTSIVAVNG 660
DB 594 LG--AKGPLYKYPDVQNECRPCHEHNCCTQCKPELQDCLGQTLVLIGKHLTMALVING 651
QY 661 ILVVVLGVFGILIKRRQOKIR-KYTMRLLOETELVPTLPSGAMPQAOIRILKETE 719
DB 652 --LVVIFMMLGSTFLYMGGRITQNRARARRYLERGESIEPLDPS-EKANKVLARIKETE 708
QY 720 LRKVKVLSGAGFYTKGIWIPDGENVKI PVAIKYLRENTSFKANKEILDEAYVMAGVS 779
DB 709 LRKLVLGSGVFGVTHKGWVMEGESIKIPVCIKYIEDKSGRQSFQAVTDHMLAIGLDH 768
QY 780 PYSRLIGICLSTVQVLTQMLPVGCLLDHRENRGRLSQDOLLWCMQIAGKMYLEEDV 839
DB 769 AHIVRLGLCPRESSIQVLTQYPLGSLDHDVHQHGAUGPOLLMWGVQIAGMYLLEH 828
QY 840 RLVRHDLAARVLYKSPHNVKITDPGLARLIDIDETERYHADGKVPKIMMALESILRRF 899
DB 829 GMVHKNLARVLYLKSPOVOVADRGVADLPRDDKOLLYSAKTPIKMALESIFGKY 888
QY 900 THQSDVMSYGVTTWELMTFGAKPYDGI PAREIPDLLEKGERLPPRPICITIDVYMIWVKM 959
DB 889 THQSDVMSYGVTTWELMTFGAEPYAGRLAEVPDLLEKGERLAQOICTIDVYMWVKM 948
QY 960 MIDSCRRPRELVSFESRMRADPORFVIONEDGPA---SPLDSTFRSLLEDDMD 1016
DB 949 MIDENIRTFEELAEFRMARDPPRYLVIKRES-GFGIAPRPEHGLTNKKLEVELEPR 1007
QY 1017 LVDAEYLVLPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTGLER-SEEAAPRSP 1075
DB 1008 ELUDLDLEAED-----NLATTLTGLSLSPVGLTNPRSGQSLL 1048
QY 1076 APSEGAGSDVFDGDMGAAKGLSLPTHD-PSPLQRYSEDPVLP-----SETDGVY 1128
DB 1049 SPSGSGY-MPMNQGNIGESQGSAAVSGSSERCPRVSLH-----PMRGLCLASSESEGHV 1101
QY 1129 A-----PLTSGPOPE---YVNPQDVAPRPPSPREGP-----L 1157
DB 1102 TGSEAELOEKVSMCRSRSRSPRPRGSAVHSQHSLLTPVPLSPGLEEEDVNGVYM 1161
QY 1158 PAARPAGATLERAKTLSP-GKNGVY-----KDVAFCGAVENPEYLTPOGGAARQPPRP 1210
DB 1162 PDLHLKGTSPSRREGTLSSVGLSVLGTBEEDD-----BEYETMNRRRRHSPP 1212
QY 1211 PAFSPAFDNLVYWD-----ODPERGAPRSPFKGPTANPEYL 1249
DB 1213 RPSLSLELGYEYMDVGSLSASLSTGSCPLHPVPIMPTAGTTPDEDEYEM 1263
```

RESULT 10  
ERB3\_RAT  
ID\_ERB3\_RAT STANDARD; PRT; 1339 AA.  
AC Q62799; Q62955;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)  
DE (C-erbB3).  
GN ERB3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=96096535; PubMed=8522190;  
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
RT "Cloning of the rat ErbB3 cDNA and characterization of the  
recombinant protein."  
RL Gene 165:279-284(1995).  
RN (2)  
RP REVISIONS TO 85; 513 AND 565.  
RA Hellyer N.J., Koland J.G.;  
RL Submitted (DEC-2001) to the EMBL/Genbank/DBD databases.  
RN (3)  
RP SEQUENCE OF 922-1097 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
RX MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neuroligin and their putative receptors, ErbB2 and  
ErbB3, is induced during Wallerian degeneration."  
RL J. Neurosci. 17:1642-1659(1997).  
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
(POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
PHOSPHATIDYLINOSITOL 3-KINASE.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL collaboration -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U29339; AAC28498.2; -  
DR EMBL: U52530; AAC53050.1; -  
DR HSSP: P11362; 1FGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_kinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001345; Tyr\_kinase.  
DR Pfam: PF00069; Kinase; 1.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR PRINTS: PR00109; TYRKINASE.  
DR Prodom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00261; FU\_5\_pkinase; 1.  
DR SMART: SM00219; TYRK; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KV Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 19  
FT CHAIN 20 1339  
FT DOMAIN 20 643  
FT TRANSMEM 644 662  
FT DOMAIN 663 1339  
FT DOMAIN 183 259  
FT DOMAIN 707 964  
FT NP\_BIND 713 721  
FT BINDING 740 740  
FT ACT\_SITE 832 832  
FT DISULFID 186 194  
FT DISULFID 190 202  
FT DISULFID 210 218  
FT DISULFID 214 226  
FT DISULFID 227 235  
BY SIMILARITY.

FT DISULFID 211 243  
FT DISULFID 246 255  
FT DISULFID 259 286  
FT DISULFID 290 301  
FT DISULFID 305 320  
FT DISULFID 323 327  
FT DISULFID 500 509  
FT DISULFID 504 517  
FT DISULFID 520 529  
FT DISULFID 533 549  
FT DISULFID 556 573  
FT DISULFID 576 585  
FT DISULFID 589 610  
FT DISULFID 613 621  
FT DISULFID 617 629  
FT CARBOHYD 126 126  
FT CARBOHYD 250 250  
FT CARBOHYD 353 353  
FT CARBOHYD 408 408  
FT CARBOHYD 414 414  
FT CARBOHYD 437 437  
FT CARBOHYD 469 469  
FT CARBOHYD 522 522  
FT CARBOHYD 566 566  
FT CARBOHYD 616 616  
FT CONFLICT 1028 1028  
SQ SEQUENCE 1339 AA; 147545 MW; 0AASF2402BBFD1E CRC64;  
Query Match 33.6%; Score 2283.5; DB 1; Length 1339;  
Best Local Similarity 40.4%; Pred. No. 3.2e-116;  
Matches 518; Conservative 170; Mismatches 440; Indels 155; Gaps 34;  
QY 3 LAALCRKGLLALLPPGAA---STVCTGTDMKRLPLSPSPETHLMDLRHLVGGCCVQGN 59  
DB 7 LQVLC---FLTSLARGSEMNSQAVCPGTLNGLSVTGDADQYQTLKLYLKECEVMGN 62  
QY 60 LETLTPTNASSIFLDIOEVQGYLIHANOVRQVPLRLRIVRGTQLFEDNYALAVDN 119  
DB 63 LEIVTGNADSLFQIMREVTGYLVAMNESVPLRLRIKRVKGTQYVDGFAIPVM-- 120  
QY 120 GDPLNNTTPVTGASPGGLREIQLRSLEILKGVLIQRNPOLCYODTILMKDIPIKNNOL 179  
DB 121 ---LNYNT-----NSHALRLQKFTQLTEILSGVYIENKDKLCHMDITDMRDIIVVR-- 170  
QY 180 ALTILDTNRSRACHCSMPCKSRGWSSSECCSLTRTVCGGC-ARCKGLPQYIKAN 238  
DB 171 GAELIVKXNNGANCPCEHVCCKG-RCMGPGPPDCCQILITKICAPCCNGRCFGPNPQCCHD 229  
QY 239 SKFIGITELKISDCLACLPFNHSGICELHCPALVYNTDTPFSMNPBEGRYFGASCYTA 298  
DB 230 ECAGSGSPQDITDCACRFNDSCGVCPRCEPLVYNNKLTLPLEBPHTKYQYGGVCVAS 289  
QY 299 CPYNTLSTDVSGCTLVCEPLHNOEVTAEQDTCRCEKSKRCARVCYGLGMEHLREVAVTS 358  
DB 290 CPNHFV-VDQFFCVACPPDKMEVD-KHGLKMKCEPCGGLCPVACGCTGSG--SRQYTVDS 345  
QY 359 ANIOFACCKTIFGSLALPESFDPPASNTAPLQPELOVETLEETIGYLIYIAMPWS 418  
DB 346 SNIDFVNACTKILGLDELITGLANDPWPKIPLDPEKLNVRRTYREITGYLNIQSWPH 405  
QY 419 LPDLSVFNQLQVIRIRLIHNGAYS-LTLOGIGISWLGRLSELGSLALHNNHTLCV 477  
DB 406 MENEVFVSNLTTIGRSLYNKGFSLLIKKLVNTSLGRSLKEISAGRYIISANOQLCH 465  
QY 478 HTVPMDQLFNPQHALHTA-NRPBDCVGEGLACHQLCARGCHGPGPTQVCNSQFLR 536  
DB 466 HSLNMTLRLLRGPSEERLDIKYDRPLGECLAEKVCDDPLSSGGCGPGGQCLSCRNYSR 525  
QY 537 GQCEYEBRVYOGIPREVYNAHCLPCHPEQCPQWGSYTCRGPADOCYAACHYADPPC 596  
DB 526 EGVCYTHGNPLQGBERFVHAQCFSCHPCELPMEGISTCNGSGSDACARCAHFPDGPIC 585  
QY 597 VARCPGKVPDLSTYMPIMKFPDEBGACQPCPINCTHSC--VDLDDKCGPAEGRASPLTSI 654

```

Db      586 VNSCHGILG--AKGPIKYPDAQNECRPCHECTOGCGNPGELQCLGQAEVLSKPHLV 643
Qy      655 USAVGILLVVLGVFGILLIKRROOKIR-KYTRRLLOETELVPLTPSGAMPQAOQR 713
Db      644 IAVVVG--LAVIIMILGSSFLYMRORRIONKRAMRYLERGESIFPLDP--EKANKVLAR 700
Qy      714 ILKTELKRVKVLGSGAGFYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYV 773
Db      701 IFKETEIRKLKVLGSGVGTGHKGIWIEGESIKIPVCIKVIEDKSGQSFQAVTDHMLA 760
Qy      774 MAGGSPVSRLLGICLTSTVQLTQMLPVCGLDHNVENGRGSGOLLWMCQIAGM 833
Db      761 VGSIDHAHIVRLGLCPSSSLQVLTVQYPLSLDLHVQHETGLPOLLWVGQIAGM 820
Qy      834 SYLEDVRLVHRDLAARNVLVKSPTHKITDGLARLLDIDETEVADGKVPKIMMALES 893
Db      821 YLEEHSMVHRDLALRNVLKSPGVQVADFGVADLLPPDDKQLHSEAKPIKMALES 880
Qy      894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
Db      881 IHFKYTHQSDVMSYGVTVWELMTFGAPYAGLRLAELPDLLEKGERLAPQICTIDVYM 940
Qy      954 IMVACWMIIDSECRPRELVEEFSRMARDPQRFVIONEDGPSPLDSTYRSLIEDD 1013
Db      941 VMVCMWIDENIRPTELANEFTRMARDPPRYLTKAS--GPGTP--PAAEPSVLTITKE 997
Qy      1014 MGDVLDAEYLVPOOGFPCPDPAAGAGMHHRRSSSTRSGGDLTGLEPSEE----- 1068
Db      998 L-----QEAELEPEL-----DLDLDLAEERGATLS 1023
Qy      1069 -----EABRPLAPSEG-----AGSDVFDGLGMAKAGLSLPTHD 1105
Db      1024 LGSALSLPTGLTLPKRGQSLLSPSSGTMNOSLGRACLDASVLAGREGFSPRISLH- 1082
Qy      1106 PSPLOKRSBDPTVLPSETDGV---APL-----TC-----SPOE-----YVNOQDV 1145
Db      1083 PIPGR-----PASESSEGVGSELEOKVSVCSRSSRSPRPRGSAVHSQHS 1135
Qy      1146 RPOPPSPREG-----LPAAPAGATLERAKTLSP-GKNGV-----KDVPAF 1187
Db      1136 LLTFTVTLSPGLGEEDGNGVMDTHLRGASSSREGTLLSVGLSVLTGEEDBD----- 1191
Qy      1188 GGAVENPEYLTPOGGAAPQPHPP 1210
Db      1192 -----EEYEVNMRKRGRSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-DEC-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; Pubmed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550 (1994).
RN 12
RP REVISIONS.

```

```

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; Pubmed=2982499;
RA Livnen E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607 (1985).
RN 14
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; Pubmed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101 (1986).
RN 15
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; Pubmed=9882502;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoehne R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,
RA Butts K.C., Buam D.A., Butler H., Cadieu L.B., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lau X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Syvakas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
RA Wang Z.-Y., Waseeman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN 17
RP SEQUENCE OF 959-1078 FROM N.A.
RC SMRAIN=Daekwanryeong;
RX MEDLINE=85173938; Pubmed=2983232;
RA Madsworth S.C., Vincent W.S. III, Bildeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal

```



[illegible]

RP SEQUENCE FROM N.A.  
RA MEDLINE=6522822; PubMed=2988784;  
RX Nilben T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,  
RA Crittenden L.B., Raines M.A., Kung H.-J.;  
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA  
RT processing and promoter insertion result in expression of an  
RT amino-truncated EGF receptor.",  
RL Cell 41:719-726(1985).  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS  
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS  
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY  
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE  
CC c-ERBB GENE, WHICH LEADS TO c-ERBB ACTIVATION.  
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB  
CC PROTEIN.  
CC  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, M1006; AAA48763.1; ALT\_INIT.  
DR PIR, A00643; TVCHLV.  
DR PIR, B00643; TVFVLV.  
DR HSSP, P11362; IFEK.  
DR InterPro, IPR000719; Euk\_pkinase.  
DR InterPro, IPR001245; Tyr\_pkinase.  
DR Pfam, PF00069; pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom, PD000001; Euk\_pkinase; 1.  
DR SMART, SMO0219; Tyrc; 1.  
DR PROSITE, PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE, PS00109; PROTEIN KINASE TYR; 1.  
DR PROSITE, PS0011; PROTEIN KINASE DOM; 1.  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;  
KM Glycoprotein; Phosphorylation.  
FT DOMAIN 132 399  
FT NP BIND 138 146 ATP (BY SIMILARITY).  
FT BINDING 165 165 ATP (BY SIMILARITY).  
FT ACT SITE 257  
FT ACT SITE 257 BY SIMILARITY.  
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FC CRG64;

Query Match 25.8%; Score 1749.5; DB 1; Length 634;  
Best Local Similarity 52.3%; Pred. No. 1,le-87;  
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17

Gy 587 CAHYKDPFCVCARCSCGVKPDLSYPIWKFPDEBEGACCPFINCHTSCVDLDKCGPAEQ 646  
Db 3 CAHPLDGHCVAKCAGVAGVGENDTL-WMKYADANAVCOI-CHNPCTRGCKGPGLGCP-- 58  
Gy 647 RASPLSTIVSAVV-GILLVVVAGVVGIIKKRROOKIKRYTKRRLLQOETVELPTPSGA 705  
Db 59 NGSKTPSIAGA VGGLCLLVVVGIGLTYLRNR-HIVKRTLRRLQOEBELVEPLTPSGE 117  
Gy 706 MPNOAQMRILEKETELRKVKVLCGSAGFGTVYKGIWIPDGENVKIPVAIKVLENTSPANK 765  
Db 118 APNOAHRIILKETEPKRVKVLCSGAFGYVKGKLMIPEGEKVKIPIVALKELRATSPANK 177  
Gy 766 ELIDEAAYMAVGSPYVSRLIGICTLTSTYQLVTOLMPYGCLLDHVRENGRLQSODLLNW 825  
Db 178 ELIDEAAYMAVDNPHVCRLLGICTLTSTYQLTTOCMPYGCCLLDYIREHDNDIGSOYLWW 237  
Gy 826 CMQIAKGSYSLEDVYLVARBDLAARVNLVSPHNVITTDGLARLLDIDETEVHADGGKVP 885  
Db 238 CVQIAKGNMYIEERRLVARDLARVNLVKTPIPVHVIITDGLKMLGADEREYTHABGGKVP 297  
Gy 886 IKMNALESILRRRFTHOSDVMSYGVTWELMTFGAKPYDGIPAREIPTDLLEKGERLPQPP 945



Db 298 IKMALBESILRIRYTHQSDVWSYGTWVWELMTFGSKPTDGI PASISSTVLEKGERLPQPP 357  
Qy 946 ICTIDVYIMWYKCMWIDSECRPRFRELVSFSGRMARDPQRFVIO-NEDLGPASPLDSTF 1004  
Db 358 ICTIDVYIMWYKCMWIDSECRPRFRELVSFSGRMARDPQRFVIO-GDERMHLPSPTDSKF 417  
Qy 1005 YRSLIEDDMGDVNAEEYVLPQGFPCRPDPAFGAGVNHHRSSSTRSGGDLTLGLE 1064  
Db 418 YRSLIEDDMGDVNAEEYVLPQGFPCRPDPAFGAGVNHHRSSSTRSGGDLTLGLE 449  
Qy 1065 PSEERAPSPPL-----APSEAGSDVFPDGLGMAKGLQSLPTDPSPLQRYSEDPVP 1119  
Db 450 -----SRTPLLSLSATSNNNSATNCID-----RNGQGHFPAVEDSFVQRYSSDPICN 495  
Qy 1120 LPSET--DGYVAAPLTCSPOPEYVNPQDVPPOPPSPRECEPLPAARPAAGATLERAKTLP 1177  
Db 496 FLEESIDIDGFL-----PAPEYVNO--LMPKKPS-----TAMVQ 526  
Qy 1178 NGVYKDVAF-----AFGAVENPEYLTPOGGAPOPHPPAPSPAPFDNLVY 1222  
Db 527 NQIYNINISLTAISKLPMDSRVONSHSTAVDNEPYL-----NTNQSFLAKTVFESSPY 578  
Qy 1223 WDO-----DPEP-----RGAPSTFKGTPTAENPEYLTGLDVP 1254  
Db 579 WIQSGNHQIQLNDNPYQODFLPNETKPNGLLKVPAENPEYLRVAP 625  
RESULT 13  
ERBB\_AVIER STANDARD; PRT; 604 AA.  
ID ERBB\_AVIER STANDARD; PRT; 604 AA.  
AC P00535;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).  
GN V-ERBB.  
OS Avian erythroblastosis virus (strain ES4).  
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
OX NCBI\_TaxID=79685;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=H;  
RX MEDLINE=84026539; PubMed=6313229;  
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;  
RT "The erbb gene of avian erythroblastosis virus is a member of the src  
RT gene family.";  
RL Cell 35:71-78 (1983).  
RN (2)  
RP SEQUENCE OF 1-152 FROM N.A.  
RX MEDLINE=84223957; PubMed=6328658;  
RA Debutre B., Henry C., Benajissa M., Biserte G., Claverie J.-M.,  
RA Sauls S., Martin P., Stehelin D.;  
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a  
RT new type of oncogene.";  
RL Science 224:1456-1459 (1984).  
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -I- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND  
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS  
CC IN CHICKENS.  
CC -I- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE  
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.  
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL, outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as there is no way in which  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: K02006; AAA42394.1; ALT\_INT.  
DR EMBL: K01216; AAA42400.1; --  
DR PIR: A00644; TVYTH.  
DR HSSP: P11362; IFGK.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00219; Tyrc; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;  
KW Glycoprotein; Phosphorylation  
FT DOMAIN 132 399  
FT NP\_BIND 138 146 ATP (BY SIMILARITY).  
FT BINDING 165 165 ATP (BY SIMILARITY).  
FT ACT\_SITE 257 257 BY SIMILARITY.  
FT CONFLICT 29 29 R -> W (IN REF. 2).  
FT CONFLICT 140 140 S -> F (IN REF. 2).  
FT CONFLICT 146 146 I -> V (IN REF. 2).  
SQ SEQUENCE 604 AA; 67653 MW; 76EBDD067450609 CRC64;  
Query Match 25.1%; Score 1703; DB 1; Length 604;  
Best Local Similarity 52.2%; Pred. No. 3,3e-85;  
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;  
Qy 587 CAHYKDPFCVACRCSGVKPDLSYMPWKFPDEGACQPCPINCTHSCVDLDDKCCPAQ 646  
Db 3 CAHFIDGPHCVACRCAVGLGENDTL-VKVDANAVCQCHNCRCGKGPGLGECPP 58  
Qy 647 RASPLTISVSAVY-GILVVYLVGVFGILKRRQKIKRTYTRRLQELVELVPLTPSQA 705  
Db 59 NGSKTPSIAAGVGGLLLVVGLIGLYLRRL-HIVKRTLRRLQELVELVPLTPSPE 117  
Qy 706 MPNQKMRILKETELRKTVLGSAGFTVYKGIWIPDGENVKIPAIKIVLRENTSPKAK 765  
Db 118 APNQHLLILKETEKTKKVLGSAGFTIYKGLWIPDEKVKIPAIKELRATSPKAK 177  
Qy 766 EILDEAYVMAGVSPYVSRLLGICLTSTVQVLVQLMPYGCLLDHVRENKRGSLDLNW 825  
Db 178 EILDEAYVMAGVSNVHVCRLGICLTSTVQVLVQLMPYGCLLDIYREHKNIGSOYLLNW 237  
Qy 826 CMQIAKGSYLEDVLYHRDLAARVLYKSPNHVITPQGLARLLDIETEVHADGKVP 885  
Db 238 CVQIAKGNVLYEERLVHRDLAARVLYKTPQHVKITPQGLAKLGADEKEYHAEGKVP 297  
Qy 886 IKMALBESILRRRFTHQSDVWSYGTWVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPP 945  
Db 298 IKMALBESILRIRYTHQSDVWSYGTWVWELMTFGSKPTDGI PASISSTVLEKGERLPQPP 357  
Qy 946 ICTIDVYIMWYKCMWIDSECRPRFRELVSFSGRMARDPQRFVIO-NEDLGPASPLDSTF 1004  
Db 358 ICTIDVYIMWYKCMWIDSECRPRFRELVSFSGRMARDPQRFVIO-GDERMHLPSPTDSKF 417  
Qy 1005 YRSLIEDDMGDVNAEEYVLPQGFPCRPDPAFGAGVNHHRSSSTRSGGDLTLGLE 1064  
Db 418 YRSLIEDDMGDVNAEEYVLPQGFPCRPDPAFGAGVNHHRSSSTRSGGDLTLGLE 449  
Qy 1065 PSEERAPSPPL-----APSEAGSDVFPDGLGMAKGLQSLPTDPSPLQRYSEDPVP 1119  
Db 450 -----SRTPLLSLSATSNNNSATNCID-----RNGQGHFPAVEDSFVQRYSSDPICN 495  
Qy 1120 LPSET--DGYVAAPLTCSPOPEYVNPQDVPPOPPSPRECEPLPAARPAAGATLERAKTLP 1177  
Db 496 FLEESIDIDGFL-----PAPEYVNO--LMPKKPS-----TAMVQ 524  
Qy 1178 NGVYKDVAF-----AFGAVENPEYLTPOGGAPOPHPPAPSPAPFDNLVY 1218  
Db 527 NQIYNINISLTAISKLPMDSRVONSHSTAVDNEPYL-----NTNQSFLAKTVFESSPY 574  
Qy 1219 NLVYWDODPPRPGAPSTFKGTPTAENPEY 1248



Db	178	EILDEAVVMA5VNDPHVCRLLGICLTSTVGLIQLMPEYGCGLDVIIRHKONIGSYLLNM	237
Qy	826	CMQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGSKVP	885
Db	238	CVQIAKGMNLYEERHMYHRDLAARNVLVTKPHVKITDFGLAKQGADEKEYHAEQGVKVP	297
Qy	886	IKKMAESILRRRTTHSDVWSGVYTWELMTFGAKRYDGIIPAREIFDLEKGERLPPQ	945
Db	298	IKMAALESILHRTIYTHSDVWSGVYTWELMTFTGSKRYDDIIPASEISSVLEKERLPPQ	357
Qy	946	ICTIDVYMIWKCMMIDSECRPREPRLVSEFSNRAPQRFVYIQ-NEDGASPLDSTF	1004
Db	358	ICTIDVYMIWKCMMISADSRPKFRELIAFSSKARDPPYLYIQGERMNLPSPTDSKF	417
Qy	1005	YRSLLDDMGDLVDAEYLVPOQGFPCPDPAAGAGMWHRRHSSTRSGGDLTLGLE	1064
Db	418	YRLTMEEDMEDIVDADEVLYVPHQGF-----NSPT-----	449
Qy	1065	PSEEARSP-----APSGAGSDVFDGLGCAKGLSLPTHDPSPLORYSEDPVP	1119
Db	450	-----SRPTLSSLSATSNNSATNCIDRMG-----H-----	476
Qy	1120	LPSETDGYVAPLTCSPQPEYVNOVDVAPQPSRREGPLPAAPAGAT-LEAKTLSPGN	1178
Db	477	PVREDFL-----PAPRYNQ-LMPKKPSTAMOVONQIYNSLTALSKLPIDSKYN	527
Qy	1179	GVWQDVFAFGAVENPEYL 1197	
Db	528	-----SHSTAVDNPPEYL 539	

RESULT 15

EGFR\_CHICK

ID	EGFR_CHICK	STANDARD	PRT	703 AA.
AC	P13387			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER			
GN	EGFR			
OS	Gallus gallus (Chicken)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus			
OX	NCBI_TaxID=9031:			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88261272; PubMed=3260329;			
RA	Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,			
RA	Ullrich A., Vennstrom B., Schlessinger J., Givol D.;			
RT	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,			
RT	expression in mouse cells, and differential binding of EGF and			
RT	transforming growth factor alpha.";			
RL	Mol. Cell. Biol. 8:1970-1978(1988).			
CC	-1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.			
CC	AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND			
CC	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- MISCELLANEOUS: Binding of EGF to the receptor leads to			
CC	dimerization, internalization of the EGF-receptor complex,			
CC	induction of the tyrosine kinase activity, stimulation of cell DNA			
CC	synthesis, and cell proliferation.			
CC	-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(Without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-210-224-12

Perfect score: 6789  
Sequence: 1 MELALACRGLLALLPFGA.....TRKGTPTANPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	6187	91.1	1259	6	018735
2	3079	45.4	1209	11	090X70
3	3050	44.9	1210	11	09BP98
4	2659	39.2	1165	13	09YH40
5	2635.5	38.8	1137	13	09W6F6
6	2241	33.0	1128	13	P79754
7	1961.5	26.0	1433	5	09BH19
8	1766.5	26.0	419	4	09UK79
9	1739	25.6	367	11	08R2X1
10	1720	25.3	729	15	086712
11	1718	25.3	567	15	086714
12	1697.5	25.0	412	4	08WYV0
13	1653.5	24.4	962	15	064895
14	1645	24.2	545	15	085468
15	1441.5	21.2	655	11	09WVF5
16	1425.5	21.0	643	11	09BRV6

17	1217	17.9	1193	5	09Y1X8	09Y1X8 ephydattia f
18	1161.5	17.1	1368	5	023821	023821 caenorhabdi
19	1155	17.0	1717	5	026566	026566 schistosoma
20	1063	15.7	527	13	090836	090836 gallus gall
21	971.5	14.3	599	13	09PSH2	09PSH2 gallus gall
22	936.5	13.8	478	11	09RSE0	09RSE0 rattus norv
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.1	176	11	0923V5	0923V5 rattus norv
25	806.5	11.9	346	13	P11776	P11776 xiphophorus
26	778	11.5	435	5	08SZM1	08SZM1 drosophila
27	754.5	11.1	311	13	099162	099162 xiphophorus
28	712.5	10.5	1362	13	09PVZ4	09PVZ4 xenopus lae
29	683	10.1	1671	5	09NIV5	09NIV5 biophalar
30	658.5	9.7	1368	13	08UW85	08UW85 paralicthy
31	648	9.5	331	4	09BUD7	09BUD7 homo sapien
32	643	9.5	1418	13	093457	093457 scophthalmu
33	626.5	9.2	1472	5	09U5A8	09U5A8 bombyx mori
34	626	9.2	1358	13	073798	073798 xenopus lae
35	622.5	9.2	1369	13	08UW86	08UW86 paralicthy
36	615	9.1	149	6	09BG66	09BG66 oryctolagus
37	611	9.0	1412	13	08UW84	08UW84 paralicthy
38	602	8.9	987	11	091YMO	091YMO mus musculu
39	599	8.8	987	11	099MR2	099MR2 mus musculu
40	595	8.8	1371	11	09QVW4	09QVW4 rattus norv
41	592.5	8.7	2144	5	09VD94	09VD94 drosophila
42	590.5	8.7	1418	13	08UW83	08UW83 paralicthy
43	587.5	8.7	1036	4	007912	007912 homo sapien
44	587.5	8.7	1055	11	054967	054967 mus musculu
45	586.5	8.6	1091	4	09UWQ4	09UWQ4 homo sapien

## ALIGNMENTS

RESULT 1  
018735  
ID 018735  
AC 018735  
DT 01-JAN-1998 (TREMREL. 05, Created)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE EtbD-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tokota H.;  
RT "cDNA cloning of etbD-2 from canine mammary gland."  
RL Submitted (0CT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1; -  
DR HSSP; P11363; 1FCX.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Furin-kinase.  
DR InterPro; IPR002174; Furin-kinase.  
DR InterPro; IPR001245; Tyr-kinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TyKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D9C4CAD6 CRC64;

Query Match 91.1%; Score 6187; DB 6; Length 1259;  
Best Local Similarity 91.0%; Pred. No. 0;  
Matches 1147; Conservative 39; Mismatches 68; Indels 6; Gaps 2;

QY 1 MELALCRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLLDMLRLYGGCCVVGQNL 60  
DB 1 MELAMCRWGLLLALLPSGAAGTGVCTGTDMLRLPASPEHLLDMLRLYGGCCVVGQNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVGVYVLAHNOVQVPLQSLRIVRGTOLEFEDVYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVGVYVLAHNOVQVPLQSLRIVRGTOLEFEDVYALAVLDNG 120  
QY 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLEGGIPAPGAAGGJRELQRLRSITLILKGGVLIQRSPQLCHQDTILMKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHCSFWCKSGRCMGESSEDCQSLTRVVCAGGACRCGRLPOTIKANSK 240  
DB 181 LTLIDTNRSRACHCSFWCKSGRCMGESSEDCQSLTRVVCAGGACRCGRLPOTICHEQC 240  
QY 241 FIGITELKHSDDLCLAHFNHSGICEHCPALVTYNTDFEEMPNDEGRYTGASCVTACP 300  
DB 241 AAGCTGPHGSDCLAHFNHSGICEHCPALVTYNTDFEEMPNDEGRYTGASCVTSCP 300  
QY 301 YNVLSTVGSCCTVLCPLHNOEVTAEADGTQRCCKSKPCARVCGLGEMHLEBVAVTSAN 360  
DB 301 YNVLSTVGSCCTVLCPLHNOEVTAEADGTQRCCKSKPCARVCGLGEMHLEBVAVTSAN 360  
QY 361 IOEPAGCKKIFGSLAFLPESFDCPDASTAPLOPQLOVEFTLEITGYLYISAMPDLR 420  
DB 361 IOEPAGCKKIFGSLAFLPESFDCPDASTAPLOPQLOVEFTLEITGYLYISAMPDLR 420  
QY 421 DLASFONLOVIRGRILHNGAYSLTLOGLISMLGLRSREHSGSLALIHNTHCFFVTV 480  
DB 421 NLASFONLOVIRGRILHNGAYSLTLOGLISMLGLRSREHSGSLALIHNTHCFFVTV 480  
QY 481 PMDOLFNNPHQALIHNTANPREDECVGSLACHOLCARGHCGPPTQCVNCSQFLRGQEC 540  
DB 481 PMDOLFNNPHQALIHNTANPREDECVGSLACHOLCARGHCGPPTQCVNCSQFLRGQEC 540  
QY 541 VEEGRVIOGLPREYVNAHCLPCHPECOPONGSVTTCREPADOCVACHYDDPPFCVARC 600  
DB 541 VEEGRVIOGLPREYVNAHCLPCHPECOPONGSVTTCREPADOCVACHYDDPPFCVARC 600  
QY 600 PSQVXPLDSEMPIKKFADEBGTQCPPICTHSCADDEKCPAEDQASPTSTIIAAYVG 659  
DB 600 PSQVXPLDSEMPIKKFADEBGTQCPPICTHSCADDEKCPAEDQASPTSTIIAAYVG 659  
QY 661 ILLVVLGVVFGILLIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNOAMRIKETEL 720  
DB 661 ILLVVLGVVFGILLIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNOAMRIKETEL 720  
QY 721 RKVAVLDSGARFGTYKGIWIPDGENVKIPVAIKYLRNTSPKAKELIDELVYMWAGVSP 780  
DB 721 RKVAVLDSGARFGTYKGIWIPDGENVKIPVAIKYLRNTSPKAKELIDELVYMWAGVSP 780  
QY 780 YVSLRLGICLSTVQVLTQMLPYGCLLDHVENRGLSGODLNMCMQIAKMSYLEDDR 840  
DB 780 YVSLRLGICLSTVQVLTQMLPYGCLLDHVENRGLSGODLNMCMQIAKMSYLEDDR 840  
QY 841 LVHRDLAARVNVKSPVHVKITDGLARLDIDETEHADGKVPKIMMALESILRRFT 900  
DB 841 LVHRDLAARVNVKSPVHVKITDGLARLDIDETEHADGKVPKIMMALESILRRFT 900  
QY 901 HOSDVMASGYTVVWELMTFGAKPYDGIPIAREIPDLLEGEERLPOPICTIDVYMTVMCM 960  
DB 901 HOSDVMASGYTVVWELMTFGAKPYDGIPIAREIPDLLEGEERLPOPICTIDVYMTVMCM 960  
QY 961 IDSECRPRFRELVEFSRMAADPQRFVYIQNEDLGASPLDSTFYRSLDEDDMDGLVDA 1020  
DB 961 IDSECRPRFRELVEFSRMAADPQRFVYIQNEDLGASPLDSTFYRSLDEDDMDGLVDA 1020  
QY 1021 EFLYLVPOGGFFCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080

DB 1020 EFLYLVPOGGFFCPDPAAGAGTAHRRSSSTRNGGELTLGLEPSEBEAPRSLAPSEG 1079  
QY 1081 AGSDVFDGDLGMGAAGKAGLSLPTHDPSFLQRYSDPTVPLSESDGYAALITCSQPERYV 1140  
DB 1080 AGSDVFDGDLGMGAAGKAGLSLPSQDPSFLQRYSDPTVPLSESDGYAALITCSQPERYV 1139  
QY 1141 NQPDVRPOPSPREGRPLAARPAATLER-----AKTILSPKNGVVKDVFAGAVENPE 1195  
DB 1140 NQPDVRPOPPLALGCPPLPSPPACATLERPTLSPKILSPKNGVVKDVFAGAVENPE 1199  
QY 1196 YLTPQGAAPQPHPPAFSPADNLVYWDQDPPERGAPSTFKGTPTAENBEYLGLDVPV 1255  
DB 1200 YLARGRAAPQPHPPAFSPADNLVYWDQDPPERGAPSTFKGTPTAENBEYLGLDVPV 1259

## RESULT 2

Q90X70 PRELIMINARY; PRT; 1209 AA.  
ID Q90X70;  
AC Q90X70;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Epidermal growth factor receptor.  
GN EGFR.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RX MEDLINE=90258688; PubMed=2342466.  
RA Pech L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
RA Earp H.S.;  
RT "A truncated, secreted form of the epidermal growth factor receptor is  
RT encoded by an alternatively spliced transcript in normal rat tissue."  
RL Mol. Cell. Biol. 10:2973-2982(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RA Pech L.A.;  
RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RX Guttridge K., Dawson T.L., Earp H.S.;  
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; M37394; AAF14008.1; -  
DR HSSP; P11362; IFGK.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PR01030; Recep\_L\_domain; 2.  
DR PRINTS; PR01030; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; Fc; 3.  
DR SMART; SM00219; TYRKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1209 AA; 134891 MW; 96FE87F6CC18773 CRC64;

Query Match 45.4%; Score 3079; DB 11; Length 1209;  
Best Local Similarity 49.5%; Pred. No. 1,4e-224;  
Matches 632; Conservative 166; Mismatches 365; Indels 114; Gaps 26;

QY 3 LAALCRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLLDMLRLYGGCCVVGQNL 61

```

Db      45 LAAALCAAG-----GALEEKVCOGTSNRLLTQGLFEDHPLSLQRMFNCEVZLGNLE 66
Qy      62 LTYLPTNASLSFLQDIQEVGVYVLIANQVRQVPLQRLAIRVGTQVLFEDNYALAVLNDG 121
      67 IYVQNNYDLSFLKTIQEVAGVYLIAMNTERI:PLENLQIIRGNALYENTYALAVLSN-- 124
Qy      122 PLNNTTPVVGASGGRLEQLRLSTLEILKGGVLIQRNPOLCYODDTILMKQIFHKNOAL 181
      125 -----YGVNTKTYRELPMRNLOELIGAVRFSNNPILCNMTTIQWRIV- QDVFSLN 175
Qy      182 TLIDTNRS-RACHPCSPMCKSGSRWGESSEDCOSLTRTVGAGCA-RCKGPLQYIKANS 239
      176 MEMDVQRHLTGCRKCPSCNCGSGRGEENCOKLTKIICAQCCSRRCRERSDCCCHQ 235
Db      240 KPIGITELXSDCLACLFHNHSGICELHCPALVTYNTDFESMPNEGRYTFGASCVTAC 239
      236 CAAGCTGPRESDCLVCHRFDEATCKDTCPMLLYNPPTYQMDVNPBGKYSFGATCYKCK 295
Qy      300 PYNVLTSDVGSCTLVPLRNLNQEVTAEDGTORCEKSKPCARVVCYGLGMEHLREVAVTSA 359
      296 PKNVYVTDHSGSCVRACGPDYEV-EEGVSKCKCGPCPKVCNGIGIGFKDTLSINAT 354
Qy      360 NIOEFAGCKKIFGSLAFLPESPFGDPAASNTAPLOPOLQVFETLEITGVLYISAMPDSL 419
      355 NIKHFKYCTAISGDHLILPVAFKGDSFTRTPDPREBELIKTVKXELTGFLLIQAMPENM 414
Qy      420 PDLVYFQNLQVIRGRILHNGAVSLTQGLGISWLGRLSRLSELQGLALIHNTHLGCVHT 479
      415 TDLHAFENLEIIRGRTKHQGFSLAVVGNITSLGRSLKEISDGVYIISGNRLCYANT 474
Qy      480 VWMDOLEFRPHQALHTANPREBECVGEGLAQHQLCARHGCMQPGPTQCYNGCQFIRGOE 539
      475 IMWKLEFGTPNQKTKIMNRAEKDCATNHNVCNPLSSECCWGPBETDVCVSCNVSRGHE 534
Qy      540 CVEECRVLQGLPREYVNAHRLCPCHRECOPONGSVTCFGEAQQCAVAYXDPRCVVAR 539
      535 CYDKCNILLEGREFREPNSECTIOCHPECPLPTMANITCTGRGPNCKICAKAYVNGPCVKT 594
Qy      600 CPSSGVKPDLSYMPIMKFPDEBEGACQPCPINCTHSCVDLDKCGPABORASP-LTSIVSAV 658
      595 CPSSGINGENNTL-VMKFADANNVCHLCHANCTYGACGPGIKGC--QQPEKPIPSIATGI 651
Qy      659 VGLILVNVVGVVFGI-LIKRQOKIKYTMRLLOTELVLPITBSGAMPNOQMRLIKE 717
      652 VGGLEFIV-VALGIGLFMRRLQVYKRLRLRLQRELEVPITPSGEBAPNOAHLRLIKE 710
Qy      718 TELRKVKVLGSGAGFTVYKGIWIPDGENKIPVAIVLRENTSPKANKELLDEAYVMAGY 777
      711 TEFKKIKVLGSGAGFTVYKGIWIPSEKXKIPVAITELREATSPKANKELLDEAYVMASV 770
Db      778 GSPVYSRLIGICLTVSTQVLTQMLPYGCLLDHVRENRGRLGSDLLNMCQIAKGSYLE 837
      771 DNPVHCRLLGICLTVSTQVLTQMLPYGCLLDVREHKDNIQGYLLNMCQIAKGNVYLE 830
Qy      838 DVLRLVHRLDAARNVLYKSPNHVKITDPGLARLLDIDETVHADGCVPIKMALESILRR 897
      831 DRLVLRHDLAARNVLYKTPQHVKITDPGLAKLIGAEKEYHAEGGCVPIKMALESILRR 890
Qy      898 RFTHOSDVMSYGVTVWELMTFGAKPYDGI:PAEIPDLLEKGERLPOPICTIDVYIMWK 957
      891 IYTHOSDVMSYGVTVWELMTFGSKPYDGI:PAEISISILEKGERLPOPICTIDVYIMWK 950
Qy      958 CMMIDSECRPRELIVSEFSRNARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDMD 1016
      951 CMMIDADRPRKRELILEFSKMARBPQRYLVIQGBERMLPSPDTBENFRALMEEDMD 1010
Qy      1017 LVDABEYVLPQCGFCPPDAPAGAGMHHRRHSSSTRSGGDLTLGLEPSEEDAPSPPLA 1076
      1011 VVDABEYVLPQCGFF-----NSPST-----SRTPPL 1036
Qy      1077 PSEGASDVFDGLDGMGAAGLQSLPTHDPSPLORESDDPTVLPSET--DGVVAPLTCS 1134
      1037 SLSANSN-----SSTVACINRNGSCRVADAFLOKYSISDPTSVLTEDNIDTFL----- 1086

```

```

Qy      1135 POBEYVQNPDPVRQPPSPREGPLPAARPAQATLERAKTSLSPGKGVKVDVAFGCAVENP 1194
      1087 PVEYINQ-SVPRKPAAGSVGNPVYHQPLRP-----ARGDHLHQN--PHSNANSP 1135
Db      1195 EYL-TPQGAAPPHPPAPSPAFDNLVYWDQ-----DP-----PERGAPSTF 1237
      1136 EYLNIAQ-----FTCLSSGFDSSALWIQGSHQMSLDNDPDYQDFFPKKAPNGIF 1186
Qy      1238 KGTPTAENPEYGLDVP 1254
      1187 KG-PTAENAEYLRVAP 1202
Db
Qy
Db
RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bpidermal growth factor receptor isoform 1.
GN BGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balabanramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibie N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Scheel C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balabanramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibie N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MG1:95294; Egfr.
DR InterPro; IPR000345; GYC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Btk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PR01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Btk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```



SEQ SEQUENCE 1165 AA; 129614 MW; 7f7EB38D8771A74E CRC64;  
 Query Match 39.2%; Score 2659; DB 13; Length 1165;  
 Best Local Similarity 45.2%; Pred. No. 1e-192;  
 Matches 575; Conservative 162; Mismatches 395; Indels 140; Gaps 29;

1 MELAALCRKGLALLALPPG-AAST---QVCTGDMKRLRLPASEPTLDMKRLHYOCQV 55  
 4 LELEL-----LALLLLSLIGRCSTDPDRKYCCGTSNQMTM---LDNHLYLKKKKKYYSCNV 56  
 56 VQGNLELTYLPTNASLSFLDIOEVGVYLIANQVQVPLRLRYRGOLFEDNYALA 115  
 57 VLELELTYTOENODLSFLOSIOEVGGIVLIANNEVSTPLVNLRLRGONLWYEGNFTLL 116  
 116 VLDNCPPLNNTPTVPGSPGLRELQLRSLTEILKGVLIORNPOLCYODITLWKDIFHK 175  
 117 VMENYQK-NPSSP--DYVQYGLKQLQLSNLTETLSGGVKKVSHNLLCNVEITNMWDIVDK 173  
 176 NNOLATLIDITNRSRACHPCSPMKSGRCMGSESEDCQSLTRIVCAGGC-ARCKGPLPY 234  
 174 TSNPTNMLIPHAERQCKDCPCGVNGSCVAPRGHCQKFTKLLCAECQNRRCRGPXPD 233  
 235 IKANSKRTIGITELKHSCLCLHFNHSGICELHCPALVTYNTDFEEMPNREGRTGAS 294  
 234 CNEHCKAGGCTGPRAIDCLACRDPNDGCTKOTCPPKYIDIVSHQVNDENIRYTGAA 293  
 295 CVTACPYNYLSTVSGSTLVCPPLHNOEVTADGTORCEKSKPCARVCYGLMEHLBVR 354  
 294 CVAECPNYYVTE-GACVRSKCSAGMLEVD-ENGKRSKPCDGVCPKYCDGIGISLNTI 351  
 355 AVTSANIQEPACCKITGSLAFLEPESPDGPASTAPLOPQLOVFETLEITGYLYISA 414  
 352 AVNSTNGSFNSCKINGDIIILNNSFECDPPHYKIGMDENHNLNLTVEITGYLYIMW 411  
 415 WPSLPLDLSFQNLQVIRGRILHNGAYS-LTLOGLGSMWGLRSLRGLSLAIHINTH 473  
 412 WPNMTSLSVQNLGRIIRGTTTSRGFSFVVQVSHQMLGSLSKVSAGNVLKTPQ 471  
 474 LCEVHTVPMQDLFRNPHQALHTANRPEDECEVGEGLACHOLCARGHMGHPPTQVNCQ 533  
 472 LRYASTINMRRLFRSEDSQIEYDART-----ENQCNNECSDDGCMGPFIMCVSCLH 524  
 534 FLRGQEVCECRVYQGLPREVYANRHLPCHPCEQOPONGSVTCGPRAOQVCAHYKDP 593  
 525 VDRGGRVASCNLLQGEPRAOVDGRVCVCHOECLVQDTSITCYGPGPANCSCAKHPQDG 584  
 594 PCFVACPSGVKDPDLSMPLMKRPDEGACOPCINCTHSCVDLDDGCAEQASPLTS 653  
 585 POCIPRCPHGLGDDTL-IMKYADKMGQCPCHQNCCTQGSGFSGSCGCD-IVSHSL 642  
 654 IVSAVQIILVVVLGVFGLIKRQOKIRKYTRRLLOETELVEPLTPSGAMPNQAMR 713  
 643 AVGLVSGSLITIVALLIIVLRRRRRIK-RKRTIRRLQKEKELVEPLTPSGAPNQAFR 701  
 714 ILKETELRKVKVLGSGAFGVYKGIWIPDEBNKIPVAILVLRNTSPKANKETLDEAVY 773  
 702 ILKETEFKXRVLVGSGAFGVYKGLMNPDENIRIPVAILKREATSPKYNQVLEDAVY 761  
 774 MAGVSPYVSRILGICLSTVQVLTOLMPYGCILDHRENRGRGLSGDILNMCQIAKGM 833  
 762 MASVDHHPVRLGICLTSVQVLTOLMPYGCILDYRQHOERICGQMLNMCQIAKGM 821  
 834 SYLEDAVLVRDLAARNVLVKSPPNHYKIDFGIARLLDIDETEXHAQGVPIKMALES 893  
 822 NYLEERHLVARDLAARNVLKPNHYKITFGISKLITABEKEXQAQGVPIKMALES 881  
 894 ILRRRTTHOSDWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYM 953  
 882 ILQMTYTHOSDWSYGVTTWELMTFGSKPYDGIIPAKIASVLENGERLPOPICTIEVYM 941  
 954 IMVKCMNIDSECPREFELVSEFSRMAPRQPFVJONEDLGASPLDSTFYSLLEDD 1013  
 942 IILKCMNIDSSRPRELVEGFSQMAPDSRYLVQI---NLSPSPDRRLFRSLSSDD 998

1014 MGLVDABEYLVPQGGFCPPDPAPAGAMVHRRSSSTRSGDGLTLGLEPSEBAPRS 1073  
 999 --DVYDADEVTL-----RXYRIN-RQGS-----E 1019  
 1074 PLASEGAGSDVDFGDLGMCAKGLQSLPTHPBPLORYSDPYV-PLPSTDGVAFLT 1132  
 1020 PCIPPNCH-----PVENSIALRYISDPTONALEKLDGH----- 1054  
 1133 CSPQREYVNPDPVAPQ-----PSPRE-----GLP-ARPPAGATLERAKTSLSPKNG 1179  
 1055 -----EYVNOGSESSRLSDIYNPNYEDLTDGKGPVSLSSQEAETNFSREYLVNTONS 1109  
 1180 VVKDVFAPGAVENPEYLTPOGGAAPORHPREAPSPAFDNLVYWDOPPERGAPRSTFGK 1239  
 1110 L---PLVSSGSMDDPDV---QAG-----YQAAV-----LPQTGALTGNGMF 1144  
 1240 TPTANPEYIAGL 1251  
 1145 LPAENLELYLGL 1156

## RESULT 5

Q9W6F6 PRELIMINARY; PRT; 1137 AA.  
 ID Q9W6F6  
 AC Q9W6F6  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase (fragment).  
 GN ERBB4.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OK NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDBRAIN;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.;  
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in  
 RT embryonic chick hindbrain.";  
 RL Mol. Cell. Neurosci. 13:237-258 (1999).  
 DR EMBL; AF121963; AAD31764.1; -;  
 DR HSRP; p11362; IRGK.  
 DR InetPro; IPR000494; EGFR\_L\_domain.  
 DR InetPro; IPR000719; Euk\_pkinase.  
 DR InetPro; IPR002174; Furin-like.  
 DR InetPro; IPR001368; TNFR\_c6.  
 DR InetPro; IPR001245; Tyr\_kinase.  
 DR InetPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 1.  
 DR Pfam; PF02757; YLP\_2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
 KM Kinase; Tyrosine-protein kinase.  
 FT NON TER 1  
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 38.8%; Score 2635.5; DB 13; Length 1137;  
 Best Local Similarity 46.5%; Pred. No. 5.9e-191;  
 Matches 530; Conservative 167; Mismatches 365; Indels 77; Gaps 24;

161 LCYQDTLWKDIFHKNNQALTLIDITNRSRACHPCSPMKSGRCWGESEDCQSLTRTVVC 220

```
DB 3 LCPADITHMODIVNRPMASNFTLVPTNGSSGGRCHKSCGTG-RCMGPTENHCQTLTKTVAC 61
QY 221 AGGC-ARGKGPLPOYIKANSKFIGITELKSHDCLALCPHNSGIELHCPALVYNTDTF 279
DB 62 AEQCDGRCYGVYSDCCRRREGACGSGSPKDTDCFCACMNFNSGACVTCQPPQTFVNPPTTF 121
QY 280 ESMNPBGRYFGASCVTACPNYLTSDVSGCTLVCPHNOEVTAEADGOREKSKCA 339
DB 122 QLENNHNAKTYGAFVCKKCHNFV-VDSSSCVACBSSKHEV-BENGIKCKPCTDTCR 179
QY 340 RVCYGLGMEHLREVAVTSANIIOEPAGCKIFGSLAPLPESPFGDPASNTAPLOEOLQV 399
DB 180 KACQIGTGLSVSAQTVDSNIDKFINCTKINGNLIFLVTHGHDGPHYTHIAINEXKINI 239
QY 400 FETLEITGCVLYISAMPDLSLDLSFQULQVIRGILNNGAVSLTLQGLGSLMLBRLR 459
DB 240 FQVAREITGYLINTOSWPNMTDFRVFSLVYTGALYSGSLLLKQOQITSLQFOSLK 299
QY 460 ELGSGALALIHNTHLCPVHTVPMQDLPNNPHQALHTANRPDECEGEGACLOLCARGH 519
DB 300 QISAGNITYTDSNLCYHTNMTSLFSTPSQKTYIHNNKKAENCTADGMCVNCNELCSSDG 359
QY 520 CMGPGPTQVNCQOFLRGQECVEBECVLOGLPREVYNAHCLPCHPEQOP-ONGSVTCFG 578
DB 360 CMGPGPDQCLSKCRFIRGRTCTESCNLYDGEFREFPANGSVCEQDPOCEKMDNMITCYG 419
QY 579 PEAOQVACAHYKOPFCVACRPSGVKPDLSMPIMKRPDEGAQOPCPINTHSCVULD 638
DB 420 GPHRCHTKCFHKGPNCEVKCPDLOQANSE--FFKYADEBRECHPHCPNCTOCCRPA 477
QY 639 DKGC-----PAEQRASPLTSIVSAVY-GILLVVLGVNFGILIKRQOKIRKYT 686
DB 478 SHDCIYPTWQSTLPQHAR--TPL--IAAGVIGGLPIIVIMKLTAVYVRKRSIK-KGRA 533
QY 687 MRLLQETELVEPLTPSGAMPNOAQMRILKTELKRVKVLGSGAGFYVKGIMWIDGENV 746
DB 534 LRRLF-ETELVEPLTPSGTAPNOAQLRIKETELKRVKVLGSGAGFYVKGIMWIDGENV 592
QY 747 KIPVAIKYLRRENTSPKANKELIDELAYVMAVGSPVVSLLGLCLSTYQVLTQMLPYCCL 806
DB 593 KIPVAIKYLRRENTSPKANKELIDELAYVMAVGSPVVSLLGLCLSTYQVLTQMLPYCCL 806
QY 807 LDHVENRGRGLSGODLNMCMQIAKMSYLEEDVLVHNRDLAARNLVSPNNHKTITDGL 866
DB 653 LDYVHEHNDNIGSOLLNMCVQIAKGMVLEBRRLVHNDLAAARNLVSPNNHKTITDGL 712
QY 867 ARLLDIDETEHADGCKVPYKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 926
DB 713 ARLLGDEKEYNADGCKMPYKMALECIHYRFTHQSDVMSYGVTVWELMTFGAKPYDGI 772
QY 927 PAREIPLDLEKGERLPOPICTIDYVMIMVCMMDISCRPRERLVSEFSMADPQRF 986
DB 773 FTRELPLDLEKGERLPOPICTIDYVMIMVCMMDISCRPRERLVSEFSMADPQRF 986
QY 987 VVIQNEI-LGPASPLDSTFYRSLLEDMDMDGLVDAEYLVPOQGFCDPAPAGAGMWH 1045
DB 833 LVIGQDDMKLPSPDSCFQNLDEEDLEDMDEEVLV-QAANIPIPIYTSSTRIDS 891
QY 1046 RHRSSTRSGGDLTLGLEPSEBEAPRS--PLAP-SEGASDVFDGLGMAKAGLOSLP 1102
DB 892 NRNOFVVRDGGYAAEQGV-PMPYRAPGCIIPAPAPQAQATAIEFDTCNGTLRQVATL 950
QY 1103 THDSPLORYSEDPVLPVS-----ETDGVVAFLTQSPQEVYNOGDVAPQPSREG 1155
DB 951 AKESSTORYSADPLVFIPEVYIRGELDEDEGTMTPMRDKPKTDYINPVEENPFVSRKNG 1010
QY 1156 PLPAP-REPAATLERAKTLSPGKNGVVDV-----AFGAVENPEYITPOGAPO 1206
DB 1011 DLQAVDNDEYHN-----AFNGQPKADEYVNEPLVANTFANTLENLEYL-----K 1055
QY 1207 PHRRPAPSPADNLYWDPDEREGA--PSTFKCTPT-----AENPEYL 1249
```

```
DB 1056 NNLEPKAKKAFDNPDMYHNSLPFRSTLQHPDVLQESYTKFYKNGRIRPIVAENPEYL 1114
RESULT 6
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TREMBLrel. 03. Created)
DT 01-MAY-1997 (TREMBLrel. 03. Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21. Last annotation update)
DE Erib3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=1077531;
RA Gellner K., Brenner S.,
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes."
RT Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Prodom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;
Query Match 33.0%; Score 2241; DB 13; Length 1328;
Best Local Similarity 39.5%; Pred. No. 6-6e-161;
Matches 515; Conservative 153; Mismatches 411; Indels 224; Gaps 32;
9 WGLIALLLP--GAASD-----VCTGDMKLRLPASRETHDMLRHLYOGCVOVGNTLE 62
DB 4 WRLIMCVASRLRAASSQTOEAVCPGTQNGLSSTGSOENYVNLNDRYKGCETIMGNDLEI 63
QY 63 TYLPTNASTFLQDIOEVQGVVLTAAHNGVQVPLQRLAIVRGTOVFEEDNVALLVLDNDP 122
DB 64 TOIESNMDPSFLKTIREVTVGLVLAHNFQEIPLQQLVIRKNSLYERRFALSVFLN-- 120
QY 123 LNTTPVYVAGSPGGLREIQLSLTEILKGVLYIQNPOLCYODTILMKDIFHKNNQLALT 182
DB 121 ----YKDG--PSGLNQLGMLNLTREILDGVIQIINKKTLRGPMYWMHDIIRNNDAPIE 173
QY 183 LIDNRSAPCHPCSPMKGRKSCWGESSESDCSLRTTVGAGC----- 224
DB 174 IOFNGERGVCN--KSC-GNYCWMGPKQCOQILTKTVCAPOCNDRCFGTSPRDCCHIECA 229
QY 225 ARCGGPLPOYIKANSKFIGITELKSHDCLALCPHNSGIELHCPALVYNTDTFESMPN 284
DB 230 AGCGGPL-----DTDCFACRLFNDSSACVPOCCOTLIYNKOTFOMETN 272
QY 285 PEGRYTFGASCVTACPNYLTSDVSGCTLVCPHNOEVTAEADGTOE-CEKSKPCARVCY 343
DB 273 PNAKYQVGSICVSCQPTHFV-VDSSSCVSVPRDMEV--ERGSROCELSGCLCPKVC 329
QY 344 GLGMEHLREVAVTSANIIOEPAGCKIFGSLAPLPESPFGDPASNTAPLOEOLQVETL 403
DB 330 GTGAE--QRQTVDSNIDSFINCTKIQGSILHFLVTVGILGDDFQVNPVLAKKLEVFRTV 386
```



```

Qy 404 EETGTYLISAWPDLSDLVFONLQVIRGRILNNGAVSLTLOGLISGLRLSRLSGLS 463
Db 387 REITDILNIOGMPKELNDLSVSSSLTYTIOGRSLFRFSLMWRIPTLTLRLSRLSRLS 446
Qy 464 GLALIHNTLHCFVHTVMDLFRNPH-QALLHTNRPEDCVGBGLACHOLCARHGCMG 522
Db 447 GSVIISONAHLCYHTHTVMTOLFGRSRYRANSLNSNRMAECVADGRCDPLCSGSGCMG 506
Qy 523 PGPTQVNCQSLRGQECVECRVLQGLPREYVNAH-CLRCHPECPONGSVTCFGEA 581
Db 507 PGPOCLSCRNSYSHGTCVACGHFNSGIPREPAGLNGVCAHPECKQOTKASCTGGA 566
Qy 582 DQACAHYNDPRPCVARGCVKPDLSYMPIMKFPDEGACORPINCISCVDLDDKG 641
Db 567 DECAACRFRDGPYCMSSCPAGVN-DEGKGLIFKPNNEGCEPHONCTGCGSGPGLND 625
Qy 642 CPARQASPLTSISAVVGLLVVVLGVF-----GILKRBQOKIRKTYMRLLOET 694
Db 626 C---LEAARLTSSGQITGLALGPAGLIFCLVLFGLMLYHRLALRKAMRRYLSG 682
Qy 695 ELVEPLTPSGAMPNOQRILKTELARKVYLGSGAFGTYYKGIWIPGENVKIPVALKV 754
Db 683 ESFEPLGP-GEKGTAKVHARILKPSDLRKIKPLGSGVFGTVSKGFWIPGETVKIPVALKT 741
Qy 755 LRENTSRKANKEILDEAVTMAGVSPYVSRILGICLTSTVOLVQMLMYGCLLDHVENR 814
Db 742 IODSSGOTFTETIDHLISWGLSHLPYIVRLIGICPGCLDLVYOLSHSGSLLEIRHOK 801
Qy 815 GRLLSQDILNMCQIAGKMSYLEDVRLVHRLDIAARNVYKSPNHYKIDFGARLLDDE 874
Db 802 TSLDPQLNMCVQIAGKMYLLEHRYVHKLAAKNILKANDYOVQISDYVADLLVPLDD 861
Qy 875 TEYHADGKVPKIMMALESILRRPFTHQSDVMSYGVTVWELMTGAKPYDIPARELPDL 934
Db 862 KKYVYSEKTPKIMMALESILFRYTHQSDVMSYGVTVWEMMSGAERYASVQOGEVPSV 921
Qy 935 LEKERLPORPITCTIDYVMVWKCMIDSECRPREFVSEFSMARDPQRFVYIQNEDL 994
Db 922 LEKERLSOPACITIDYVMVWKCMIDENRPTFKELASDFTMARDPREYVLIYRMG- 980
Qy 995 GPASPLDSTFYRSLDEDDMDLVDAEYLVPOQGFPCPDAPGAGVHHRHSSSRSS 1054
Db 981 -----EDSGKREFL-----RGSER- 995
Qy 1055 GGGDLVLGLEPSEEEAPRSLAPSEAGSDVFDGLGNG---AAKGLSLPHTHPSPLQ- 1110
Db 996 --GILBADLEDEE-----GLGDRFATPISLOPSSSWETSPOI 1032
Qy 1111 -----RYSEDPYVPLPSETDGYVAPLTCSPQ- EYVNO-----PDVR 1146
Db 1033 NSYVMVMTQLRYD-----FAVSQGHIGYLPMSPSVDTIRQLWYORSRLSSVRLPDRS 1086
Qy 1147 POPSPREGPL--PAARPAATLERAKTLSPGKNGVAVDPAFGAVENPYYLPOGGA 1204
Db 1087 AFRSSSEAELECEGAQACAGIFRYR-----FSSERN-----POGG- 1122
Qy 1205 POPHPAPAFSAPFADNLVYWDODPRPERGAPSTFKGTPTAENPE 1247
Db 1123 -----OQRKLTSTASSPSSFKTMADEDE 1146

```

```

OC Anopheles.
OX NCBI_taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:SU4.
RA Lycett G.3.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ13655; CAC35008.1; -
DR HSSP: P11362; 1GCK
DR InterPro: IPR000345; CyC_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; EGFR_L_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fu_7.
DR SMART: SM00220; S_TKc; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KM Receptor.
FT NON TER
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 28.9%; Score 1961.5; DB 5; Length 1433;
Best Local Similarity 32.1%; Pred. No. 1,2e-139;
Matches 465; Conservative 196; Mismatches 401; Indels 385; Gaps 36;

Qy 26 CTGTDMKRLRLPASETHLDMRLHLYOGCOVQGNLELYLPTNASSLFLDIOEVOGYVL 85
Db 1 CIGTNGMSVAPANEHYHKLRLDRYNTCTYDGNLEITWIONITDNLQHLREVTGYVL 60
Qy 86 IAHNOVROVPLQRLIRVGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140
Db 61 ISLYDLPOVILPRQIIRGRTTFKLKKEAYGLFV-----SFSHNTL 104
Qy 141 QLRSTELIKGCVLIQGNPOLCYODTILMKOI-FHKNNQALLTLIDTRSRACHGSPMC 199
Db 105 ELPLRLDILGSGVGFENNYNLCHKMSINWEIILLAPOTSQYTFNFFSSPERVCPCHPSC 164
Qy 200 KGSRCWGESSEDCSLTRTVACAGCA--RCKGPLPOVIKANSKFIGITELKHSDCLACTH 257
Db 165 EVG-CWEGGANNCGRFSLKNCSPGSGRCGPAPRECHLFCAGGCTGPTQSDCLACKN 223
Qy 258 FNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVSGCTLVPL 317
Db 224 FYDGVCKQCECPMQIINPTNYFEPNPDGKAYAGATCVKRCP-EHLKDNAGACVAKCPK 282
Qy 318 HNGVTEADGTQREKSKPCARVCTGLGHEHLREAVNTSANTIOEPAACKKIFGSLAF 377
Db 283 GKMPDSE-----CVPKGVCPKTCPEBGIVH-----SDNIGNYDCDTIIGSLEIL 329
Qy 378 PESFDGDPASNT-----APLOPELOVETLEITGYLISAMPDLSLDLVFONLQ 429
Db 330 DQSFDFGOVYVNTNSFSPRYIKIPDLLEVFSTYKEITGFINIQAHPNFTLLNYPRL 389
Qy 430 VIRGRIILHNGAY-SLTLOGLISWLGRLSRLSGLALIHNTLHCFVHTVMDLFRN 488
Db 390 VVGGROUKENLFAVSVYIVKSLKSLKSLKRVVSGSIVILENSDLCFVEDIDMSLKKX 449
Qy 489 PHQALLHTANRPPECEGEGGLACHOLCARHGCMGPGPTQVNCQSLRGQECVECRVLQ 548
Db 450 SDHEVWVQKNRMATECHEBGEBCSEOCSSKAGCMGKGEQCLCKGVVYKGCIDLSCK--- 506

```

```

QY 549 GLPREY-VNARHCLPCHPECOPONGSVTCFGEADQVACAHYDPPFCVARC----- 601
DB 507 SLPLVSVDSKTCGDCHQEKD-----FCVGNEDNGSCMNVVDGRCAVCEPTTKAM 561
QY 602 -----SGVKPDLSTWPKFPD----- 618
DB 562 NGTCINCHKTGCGRPRDTIAPDGCISCDAKIIIGSDAKIERCLMKESCPDGYSDVYL 621
QY 619 -EEG----- 621
DB 622 QEEGRLKOLSGKAVCRKCHPRCKCTGCGFHEGFCOECTGYKKGEQCEDECPDYANEE 681
QY 622 -ACQCPRIINCT-----HSCVDL-----DD-----KGCRAEQ----- 646
DB 682 TRICLPCHQECRGCHGLGDHHECRNLKLPFGDPDYDNTPTTCVSNCPASHPRYKRPGEA 741
QY 647 -----RASPLTISVAVGILLVVLGVNGI---LTKRQOKIRKTYM 687
DB 742 GKIGRYCSADSMOSGLRIEPTQYKIVMGSMALILLCVGVGIAFVLFSRHKNKDAVKM 801
QY 688 RRLIOTELVPLTPSGAMPNOAQMRILKETELRKVKVLGSGAGTVYKGIWIPDGENVK 747
DB 802 TMLAGCEDESPRLPSNVGPNLTUKRIKEAIRGGVILGMAFGRYVKGVMPEGESVK 861
QY 748 IPAIKVLRENTSBRANKELIDEAYVMAGSPYVSRLGICTSTVOLVTQMLPYGCLL 807
DB 862 IPAIKVLRENTSBRANKELIDEAYVMAGSPYVSRLGICTSTVOLVTQMLPYGCLL 921
QY 808 DHVENGRGLSDOLLKWCQIAGKMSYLEDLVLRHLAARNLVYSPNNVKTITDECLA 867
DB 922 DYVNNKDKIGSKALLNWSQIARGMAVLEERRLRLAARNLVVOTPCSVKIKTVFGLA 981
QY 868 RLIDIDEYHADGKVPKIMMALESILRRRFTHOSDVYSGVTWELMTGAKPYDQIP 927
DB 982 KLDFDDEYVDAAGAKMPKIKWALECIRHVRFTSKSDVMAGGITWELTTGARYEVNP 1041
QY 928 AREIPDLLEKGERLPORPCTIDVYMTVMKCMIDSECRPRFRELVSFSMARDPORFV 987
DB 1042 AKDVPRELEIGHKLPQPDICSLDYVCILSCVMLDADARPTFKQLAETFAEKARDPRGYL 1101
QY 988 VIONEDLGRASRLDSTFRSLLEDDMDGLV----- 1018
DB 1102 MI-----PGKFKMLRPSYTVODEKDLIRTLAPVMAAAAAAGASNDVPSITIA 1152
QY 1019 DAEELVPOOGFPCPPDAPGAGCMVHNRSSSTRSGGDLTLGLEPSEEPARS----- 1073
DB 1153 ETDXYLQPKTRPSIMLRPSA-----VEPS-DEMPKSLRYCK 1188
QY 1074 -PLAP---SEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVA 1129
DB 1189 DPLKPRDETDGHEKEV-----GVGQIR-----LNLPLDEDDVYLM 1222
QY 1130 PLTCSPPQEVNQDVNRQPPSPREGPLAARPAQATLERAKTISFGKNGVGVKVFAGG 1189
DB 1223 P-TCOSQ---NOS-----TPG---YMLLIGVPA 1243
QY 1190 AVENPEYL-----TPGGAAPQPHPPAPAFADNLVYMDODPERGAPSTFKGT 1240
DB 1244 SVNDPEYLMGSTOAIAGLAQSMG--PHTPP-----PNTTPNGM 1280
QY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

```

RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstein..

```

```

GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF17761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FFC1BE347E2D030C CRC64;

```

```

Query Match 26.0%; Score 1766.5; DB 4; Length 419;
Best Local Similarity 86.3%; Pred. No. 1.2e-125;
Matches 340; Conservative 5; Mismatches 42; Indels 7; Gaps 2;

```

```

QY 1 MELALCRWGLLALLPPGAASVQCTGTDMKRLPASPEHLLMLRLHYGCGVQVQNL 60
DB 1 MELALCRWGLLALLPPGAASVQCTGTDMKRLPASPEHLLMLRLHYGCGVQVQNL 60
QY 61 ELTYLPTNASISFQDIQEVGVYLIANOVQVPLQRLIVRGTOU.FEDNYALAVLDNG 120
DB 61 ELTYLPTNASISFQDIQEVGVYLIANOVQVPLQRLIVRGTOU.FEDNYALAVLDNG 120
QY 121 DPLNNTPTVTCASPGELRELQRLSTELIKGVLIQRPOLCYODTILMKQIFHNQOLA 180
DB 121 DPLNNTPTVTCASPGELRELQRLSTELIKGVLIQRPOLCYODTILMKQIFHNQOLA 180
QY 121 DPLNNTPTVTCASPGELRELQRLSTELIKGVLIQRPOLCYODTILMKQIFHNQOLA 180
DB 121 DPLNNTPTVTCASPGELRELQRLSTELIKGVLIQRPOLCYODTILMKQIFHNQOLA 180
QY 181 LTLIDITRSRACHPCSPKCSKRCWGESSEDCQSLTRTVACAGGACRCKPLPOYIKANSK 240
DB 181 LTLIDITRSRACHPCSPKCSKRCWGESSEDCQSLTRTVACAGGACRCKPLPOYIKANSK 240
QY 181 LTLIDITRSRACHPCSPKCSKRCWGESSEDCQSLTRTVACAGGACRCKPLPOYIKANSK 240
DB 181 LTLIDITRSRACHPCSPKCSKRCWGESSEDCQSLTRTVACAGGACRCKPLPOYIKANSK 240
QY 241 FIGITELKHSPLCACTPHNSGICELHCPALVTYNTDFESMPNPEGYTGASCVTACP 300
DB 241 AAGCTGPHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGYTGASCVTACP 300
QY 301 YNYLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNYLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
QY 301 YNYLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNYLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
QY 361 IQEFAGCKKIFGSLAPLPESFSDPASNTAPLQ 394
DB 356 LRMQPG--PAHPVLSFLRPSWDLVSASFYSPLAP 367

```

```

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

```

DR EMBL, J0C027080; AAH27080.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;  
Query Match 25.6%; Score 1739; DB 11; Length 367;  
Best Local Similarity 88.0%; Pred. No. 1.2e-123;  
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;  
QY 889 MALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI PARELPDLEKGERLPQPPICT 948  
DB 1 MALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI PARELPDLEKGERLPQPPICT 60  
QY 949 IDVYIMVWKCMIMDSECRPRFRELVSFSRMARDPQRFVVIQNEIDLCPASPLDSTFYRSL 1008  
DB 61 IDVYIMVWKCMIMDSECRPRFRELVSFSRMARDPQRFVVIQNEIDLCPASPLDSTFYRSL 120  
QY 1009 LEEDDDMGDLVDAEYEVLPQGGFFCPDPAPGAGVGVHHRSSSTRSGGGLTGLGSESEE 1068  
DB 121 LEEDDDMGDLVDAEYEVLPQGGFFSPDPLGTGTATHRHRSSSARSGGGLTGLGSESEE 180  
QY 1069 EABRSPLAPSEGASDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYV 1128  
DB 181 EPRRSPPLAPSEGASDVFDGDLAVGVTKGLQSLSPHLSPLQRYSEDPITLPLPETDGYV 240  
QY 1129 APTTCSPQPEYVNPQPDVPPQPSPEBEGPLPAARPAGATLERAKTSLFGKGVVADVAFG 1188  
DB 241 APLACSPQPEYVNPQPEVRPQSPPLPEGPPPIRPAGATLERPKTSLFGKGVVADVAFG 300  
QY 1189 GAVENPEYLTPQGGAAPQHPAPSPAFNLLVYVODDPERKAPSTFETGTPAENPEY 1248  
DB 301 GAVENPEYLTPQGGAAPQHPAPSPAFNLLVYVODDPERKAPSTFETGTPAENPEY 360  
QY 1249 LGLDVPPV 1255  
DB 361 LGLDVPPV 367  
RESULT 10  
ID 086712 PRELIMINARY; PRT; 729 AA.  
AC 086712;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Polyprotein.  
GN POLYPROTEIN.  
OS Avian rous-associated virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
OX NCBI\_Taxid=11950;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94203659; PubMed=8152791;  
RA Venstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
RA Johnson A., Beug H.,  
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
RT different transforming capacities.";  
RL Oncogene 9:1307-1320(1994).  
DR EMBL; S69372; AAC60725.1; -  
DR HSP; P03322; 1A6S.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR004028; Retro\_M.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PFO0069; pkinase; 1.  
DR Pfam; PFO2813; Retro\_M; 1.  
DR ProDom; PDOM0001; Euk\_pkinase; 1.  
DR SMART; SMO0219; TyKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 729 AA; 80649 MW; 84DZF6914EFDID63 CRC64;

Query Match 25.3%; Score 1720; DB 15; Length 729;  
Best Local Similarity 54.8%; Pred. No. 9.5e-122;  
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;  
QY 569 PQNGSVTCFPEPADQCVACHYKDPFCVACRCSGVVPDLSYMIWKPPEEGACQPCPI 628  
DB 141 PEETAPPKTP--DHCKCAHFTIDGPHCVACPAIGENDTL-VMKYADANAVCOLCHP 197  
QY 629 NCTHSCVLDLDDKCGCAPORASPLTSIVSAV-GILVNVGVGVGILIKRQKIRRYTM 687  
DB 198 NCTRGCGPGLBEGP--NSGKTPSTIAGVVGILCLVVGGLGLTLRRR-HIVRKTL 253  
QY 688 RLLQETELVEPLTPSGAMPNOQMRLKETELRKVKVLSGAFGVYKGIWIDGENVK 747  
DB 254 RRLQRETELVEPLTPSGEANOHLRLTKTEFKVAVKLSGAFGVYKGIWIDGENVK 313  
QY 748 IPAVAILKRENTSPKANKETLDEAYVAGVSPVSLGICLTSTVOLVTOLMPYGLL 807  
DB 314 IPAVAILKRENTSPKANKETLDEAYVAGVSPVSLGICLTSTVOLVTOLMPYGLL 373  
QY 808 DHVRENRGLSGODLLNMCQIAKMSYLEDVLRHDLAARVLYKSPHHVKTTPDGLA 867  
DB 374 DYIREHKNIGSOYLNMVCQIAKGMVLEERRLVHRDLAARVLYKTPQHVKITDGLA 433  
QY 868 RLIDIDETEHADGKVPIMMALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI 927  
DB 434 KLGADBEKEXHAAGKVPIMMALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI 493  
QY 928 AREIPDLEKGERLPQPPICTIDVYIMVWKCMIMDSECRPRFRELVSFSRMARDPQRFV 987  
DB 494 ASEISSVLEKGERLPQPPICTIDVYIMVWKCMIMDSECRPRFRELVSFSRMARDPQRFV 553  
QY 988 VIO-NEDLGASPLDSTFYSLLEDMDGLVDAEYEVLPQGGFFCPDPAPGAGVYVHR 1046  
DB 554 VIQGDREMLPSPDTSKFRYTLMEEDMEDI VDAEYEVLPQGGFF----- 598  
QY 1047 HRSSSTRSGGGLTGLGSEEEAPRSP-----APSEGASDVFDGDLGMAKGLQSL 1101  
DB 599 -NSPT-----SKTPLLSSLSATSNNSATNCID-----RNGQH 631  
QY 1102 PTHDPSPLORYSEDPTVPLPSET--DGYAPLTCSPQPEYVNPQPDVPPQPSPEBGLPA 1159  
DB 632 PYVEDSFVORYSSDPTGFLFLEESIDGFL-----PAPEYVNG--LMPKKS----- 675  
QY 1160 ARPAGATLERAKTSLFGKGVVADV-----AAGAVENPEYL 1197  
DB 676 -----TAMVQVQIYNNISLTAISKLPMSRYGNSHTAVDPEYL 715  
RESULT 11  
ID 086714 PRELIMINARY; PRT; 567 AA.  
AC 086714;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE V-erbB protein (Fragment).  
GN V-ERB.  
OS Avian rous-associated virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
OX NCBI\_Taxid=11950;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94203659; PubMed=8152791;  
RA Venstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
RA Johnson A., Beug H.,  
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
RT different transforming capacities.";  
RL Oncogene 9:1307-1320(1994).  
DR EMBL; S69372; AAC60727.1; -  
DR HSP; P11362; IFGK.  
DR InterPro; IPR000719; Euk\_pkinase.

```

DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase.1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase.1.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; 609CBA7ADF725E1 CRC64;

Query Match 25.3%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 9.3e-122;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQVACAHYDOPPEVCARCPGKPDLSYPIWKEPDEEGACQPCPINCSTHSCVDL 637
DB 1 GP--DHCKCKAHFIDGPHCVACAPAGVLENDTL-VWKYADANAVCOLCHPCTRGCKGP 57
QY 638 DDGCPAEQASPLTSIVSAVY-GILLVVLGVVFGILIKRQCKIRKYTMRELLQETEL 696
DB 58 GLECCP--NGSKTPSIAAGVVGGLCLVVGGLGILYLR--HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOAMRILKETELRKVVLGSGAFGVYKGIWIPDGENVKIPVAIKYLR 756
DB 114 VEPLTPSGEAPNQHRLIKETEFKVVVLGSGAFGVYKGLMIPGEGKVIIPVAIKELR 173
QY 757 ENTSFKANKETLDAYVAGVGSVYVSRLLGICLTSTVQLVTLQMPYCGCLLDHVENKGR 816
DB 174 EATSPKANKETLDAYVAGVGSVYVSRLLGICLTSTVQLVTLQMPYCGCLLDYIREHNDN 233
QY 817 LGSODLNMCMQIKGMSYLEVDVLRHDLAARVVLVKSPPHVKITDEGLARLLDIDETE 876
DB 234 IGSQYLLNMCMQIKGMSYLEVDVLRHDLAARVVLVKSPPHVKITDEGLARLLDIDETE 293
QY 877 YHADGKVPKIMALESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLE 936
DB 294 YHADGKVPKIMALESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLE 353
QY 937 KGEPLPQPICTIVYIMVYKCMVIDSECRPRRELVSFSSMARDPQRFVYIQ-NEDLG 995
DB 354 KGEPLPQPICTIVYIMVYKCMVIDSECRPRRELVSFSSMARDPQRFVYIQ-NEDLG 413
QY 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOGFFCDDPARPAGAGVNHHRSSSTRSG 1055
DB 414 LPSTFDSKFTKTLMEEDMEDIVDAEYLVPHQGF-----NSPST-- 454
QY 1056 GGDLTLLGLEPSEEEAPRSP-----APSEAGSDVFDGLMGAKGLQSLPTHDPSPLQ 1110
DB 455 -----SRTPLSLSLATSNSNATNCID-----RNGQGHPRVDEDSFVQ 491
QY 1111 RYSEDPVPLPSEF--DGYVAPLTCSPQPEYVNPDPVPPQPSREGPLPAARPAGATLE 1168
DB 492 RYSSDPVPLPSEIDGFL-----PAPEYVNO--LMPKSPS----- 526
QY 1169 RAKTLSPKNGVNVDF-----AFGAVENPEYL 1197
DB 527 -----TAMVQNOIYNNISLTAISKLPMSRYONSHSTAVDPEYL 566

RESULT 12
Q8WYVO PRELIMINARY; PRT; 412 AA.
AC O8WYVO;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Placentalia; Primates; Carnivora; Hominoidea; Homo.

```

```

OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF18349; AA55856.1;
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; kinase.1.
DR Pfam; PF02757; YLP.2.
DR ProDom; PD000001; Euk_kinase.1.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00108; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 25.0%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 2.1e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESIIRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLEKGERLPQPICT 948
DB 1 MALESIIRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLEKGERLPQPICT 60
QY 949 IDVYIMVYKCMVIDSECRPRRELVSFSSMARDPQRFVYIQ-NEDLG 1008
DB 61 IDVYIMVYKCMVIDSECRPRRELVSFSSMARDPQRFVYIQ-NEDLG 120
QY 1009 LEDDDMDGLVDAEYLVPOGFFCDDPARPAGAGVNHHRSSSTRSGGDLTLLGLEPSEE 1068
DB 121 LEDDDMDGLVDAEYLVPOGFFCDDPARPAGAGVNHHRSSSTRSGGDLTLLGLEPSEE 180
QY 1069 EAPRSLPAPSGAGSDVFDGLMGAKGLQSLPTHDPSPLQRYSEDPVPLPSETDQYV 1128
DB 181 EAPRSLPAPSGAGSDVFDGLMGAKGLQSLPTHDPSPLQRYSEDPVPLPSETDQYV 240
QY 1129 APLTCSQPEYVNPDPVPPQPSREGPLPAARPAGATLEAKTSLGKNGVNVDFVAFG 1168
DB 241 APLTCSQPEYVNPDPVPPQPSREGPLPAARPAGATLEAKTSLGKNGVNVDFVAFG 300
QY 1189 GAVENPEYLVTPGGGAAPQ-----HPPA--FSAPFDNL 1220
DB 301 GAVENPEYLVTPGGGAALSPFTLLPSAQPSTTSITGTRTHQSGIHPAPSKGHLKQRTQST 360
QY 1221 YYMD-QDPPER-----GAPSTFKCTPTAEN 1245
DB 361 WWMTCCQCEPQGVRRSPDVSSGREGTLAGIKRKEGPPTTSRGTCARN 410

RESULT 13
Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Gag-v-erb-A-v-erb-B protein.
GN GAG-v-erb-A-v-erb-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Avian type C retroviruses.
OC NCBI_Taxid=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; Pubmed=1969616;
RA Bruskini A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B."

```

RL Oncogene 5:15-24(1990).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
DR EMBL: X52209; CA36459.1; -.  
DR EMBL: X52211; CA36459.1; JOINED.  
DR HSSP: P10828; 2NUL.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR000536; Hormone\_rec\_1lg.  
DR InterPro: IPR001723; Sterm\_n\_receptor.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR InterPro: IPR001628; Znf\_Casteroid.  
DR Pfam: PF00104; hormone\_rec; 1.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00105; zf-C4; 1.  
DR PRINTS: PR00398; STRDHOMONER.  
DR PRINTS: PR00047; STROIPINGER.  
DR ProDom: PD000001; Euk\_kinase; 1.  
DR ProDom: PD000035; Znf\_Casteroid; 1.  
DR SMART: SM00430; HOL1; 1.  
DR SMART: SM00319; Tyrc; 1.  
DR SMART: SM00399; Znf\_C4; 1.  
DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR ATP-binding; DNA-binding; Nuclear protein; Receptor;  
KM Transcription regulation; Transferase; Tyrosine-protein kinase;  
SO ZINC-finger.  
SEQUENCE 962 AA; 108320 MW; 3C5AED791EAE95CE CRC64;  
Query Match 24.4%; Score 1653.5; DB 15; Length 962;  
Best Local Similarity 51.7%; Pred. No. 1.6e-116;  
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;  
QY 541 VEEGRVAGQLPRE-VVNAH-HCLP-----CHPCCQ 568  
DB 354 IEKQESYLLAFEHYINRKINIPFWSKLMKVADLMIGAYHARSFLMKVECPPLS 413  
QY 569 PONGSVTCFGEADQCVACAHYKDPFCVACRPSGVKPDLSYMPIMKFPDEGACQCPPI 628  
DB 414 PQE-----VGP--DHCKCAHFIDGPHCVKACPAGVGENDTL-VMKADANAYCQLCHP 465  
QY 629 NCHSCVDLDKCPRAEQRASPLTSIYSAV-VGILLVVVGVFGILIKRQOKIRKXTM 687  
DB 466 NCTGCGKPGLEGCP---NGSKTPSIAGVGVGLCLVVGGLGLYLRR-HIYRKRTL 521  
QY 688 RRLQERELVEPLTPSGAMPQAOQMRILKETELRKVKVLGSGARCTYVKGIWIDGENVK 747  
DB 522 RRLQERELVEPLTPSGAMPQAOQMRILKETELRKVKVLGSGARCTYVKGIWIDGENVK 581  
QY 748 IPVAIKVLRENTSPKANKELIDEAYVMAVGSPYVSRLLGICLTSTVQLTQMLPFGCL 807  
DB 582 IPVAIKELRENTSPKANKELIDEAYVMAVSNPHVCRLLGICLTSTVQLTQMLPFGCL 641  
QY 808 DHVRENKRGISQDLNMCQIAKMSYLEDRVLVHRDLAARNVLYKSPNHVKTIDFGIA 867  
DB 642 DYIREHNDIGSQYLWMCVQIAKMSYLEDRVLVHRDLAARNVLYKTPQVKTIDFGIA 701  
QY 868 RLDDIDETVYADGKVPKIMMALESILARRFTQSDVMSYGVYVWELMTGKAPYDGP 927  
DB 702 KQLGADKEKYHAEKGPVKIMMALESILHRIYHQSDVMSYGVYVWELMTGSKPYDGP 761  
QY 928 AREIPDLKEGERLPPICTIDVYVIMVKCMIDSECRPFRELVSFSEMARDDPOFV 987  
DB 762 ASEISSVLEKGERLPPICTIDVYVIMVKCMMSGADSRPKFRELIAFESKARDPPRYL 821  
QY 988 VIO-NEDLPASPLDSTFYRSLEDDMGDLVDAEYVLPQGFPCPDPAAGAGMYHR 1046  
DB 822 VIOQDEMHLPSPYDSKYRTLMEEDEMDIVDAEYVLPQGF----- 866  
QY 1047 HRSSSTSGGDDLGLGLEPSEEARPRSPLASBEGAGSVFPGDGLGMAKAGLSLPTTPD 1106  
DB 867 -NSPST-----SRTPLLSLSATSN-----NSATKCIDRNGG- 898

QY 1107 SPLQRYSEDPVVPVPSRTDGYVAPLTCSPQEVYVNOPVROPSPSPREGPLPAARPGAT 1166  
DB 899 -----PVREDGL-----PAREYVQ--LMRKKESTANVQNIYVIST 936  
QY 1167 -LEBAKTLSPGKGVGVKDVFAFGAVENPEYL 1197  
DB 937 AISKLPMDSRVQ-----SHSTAVNDPEYL 961  
RESULT 14  
Q85468 PRELIMINARY; PRT; 545 AA.  
AC Q85468;  
DT 01-NOV-1996 (TREMBlrel. 01. Created)  
DT 01-NOV-1996 (TREMBlrel. 01. Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20. Last annotation update)  
DE Avian erythroblastosis virus (Ts34) v-erbB gene.  
OS Avian erythroblastosis virus.  
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
OX NCBI\_Taxid=11861;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88217326; PubMed=2897102;  
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;  
RT "Common site of mutation in the erbB gene of avian erythroblastosis  
virus mutants that are temperature sensitive for transformation.";  
RL Oncogene Res. 1:265-278(1987).  
DR EMBL: X06943; CA30024.1; -.  
DR HSSP: P11362; 1FGK.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR ProDom: PD000001; Euk\_kinase; 1.  
DR SMART: SM00219; Tyrc; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR ATP-binding; Transferase; Tyrosine-protein kinase.  
SO SEQUENCE 545 AA; 60899 MW; 140DC8CCA078AP4 CRC64;  
Query Match 24.2%; Score 1645; DB 15; Length 545;  
Best Local Similarity 54.9%; Pred. No. 3e-116;  
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;  
QY 578 GPEADQCVACAHYKDPFCVACRPSGVKPDLSYMPIMKFPDEGACQCPPI NCHSCVDL 637  
DB 1 GP--DHCKCAHFIDGPHCVKACPAGVGENDTL-VMKYADANAYCQLCHPCTGCKGP 57  
QY 638 DDKCPRAEQRASPLTSIYSAV-VGILLVVVGVFGILIKRQOKIRKRYTRRLQEREL 696  
DB 58 GLEGCP---NGSKTPSIAGVGVGLCLVVGGLGLYLRR-HIYRKRTLRLQEREL 113  
QY 697 VEPLTPSGAMPQAOQMRILKETELRKVKVLGSGAFYVKGIWIDGENVKIPVAIKVL 756  
DB 114 VEPLTPSGAMPQAOQMRILKETELRKVKVLGSGAFYVKGIWIDGENVKIPVAIKVL 173  
QY 757 ENTSPKANKELIDEAYVMAVGSPYVSRLLGICLTSTVQLTQMLPFGCLLDHVRENKR 816  
DB 174 EATSPKANKELIDEAYVMAVSNPHVCRLLGICLTSTVQLTQMLPFGCLLDYIREHND 233  
QY 817 LGSODLNMCMQIAKMSYLEDRVLVHRDLAARNVLYKSPNHVKTIDFGIARLLDIDETE 876  
DB 234 IGSQYLWMCVQIAKMSYLEDRVLVHRDLAARNVLYKTPQVKTIDFGIARLLDIDETE 293  
QY 877 YHADGKVPKIMMALESILARRFTQSDVMSYGVYVWELMTGKAPYDGPAREIPDLLE 936  
DB 294 YHADGKVPKIMMALESILHRIYHQSDVMSYGVYVWELMTGSKPYDGPASEISSVLE 353  
QY 937 KGERLPPICTIDVYVIMVKCMIDSECRPFRELVSFSEMARDDPOFVIO-NEDLG 995  
DB 354 KGERLPPICTIDVYVIMVKCMMSDADSRPKFRELIAFESKARDPPRYLVIOQDERMH 413

```

OY 996 PASPLDSTFYRLSLDDEDDMGDLVDAEYLVPOGFCFDPAPAGAGMYHRRSSSTRSG 1055
DB 414 LPSPTDSKFYRTLMEEDEMDIVDADEYLVPHQGF-----NSST--- 454
OY 1056 GGDULTGLEPSEEEKPRSPRL-----APSEAGSDVFDGDLGKAAGKLGSLPTHPSPRLQ 1110
DB 455 -----SRTPLSLSLATSNNSNTNCIDRNG-----H----- 481
OY 1111 RYSEDPVPLPSETDGYVAPLTCSPQEPYXNPDVPRPOPSPREBPPLPAARAGAT-LER 1169
DB 482 -----PVRDGFPL-----PAPERVNO-LMPKXSTAMVOIQIYNTSLTAISK 523
OY 1170 AKTLSPGKXGVKVDVAFAGAVENDEYL 1197
DB 524 LPMDSRYQN-----SHSTAVDNDEYL 544

RESULT 15
OY 09WVF5 PRELIMINARY; PRT; 655 AA.
AC 09WVF5;
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-NOV-1999 (Tremblrel. 21, last annotation update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danieleen A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibhe N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/Genbank/DBS databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/10J, 129/SVJ AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBS databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana K., Saito T., Okazaki Y., Gojibori T., Bono H., Kaekawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H., Kuhl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J., Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balderas R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Guetincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Matchonni L., Mashima J., Mazzarelli J., Mombaerts P., Noriose P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P., Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wyszewski A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -
DR EMBL; AF275366; AAG28047.1; -
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB33688.1; -
DR EMBL; AK004883; BAB23641.1; -
DR EMBL; AK004911; BAB23662.1; -
DR MGI; 95294; Egfr.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ
SEQUENCE 655 AA; 72906 MW; 6834063BJBC928CB CRC64;

Query Match 21.2%; Score 1441.5; DB 11; Length 655;
Best Local Similarity 42.9%; Pred. No. 1,1e-100;
Matches 274; Conservative 98; Mismatches 244; Indels 23; Gaps 7;

OY 11 LIALLPFGAA--STVCTGTDMLRLPASBETHLDMRLHYOGGVQVGNLELYLPTN 68
DB 14 LITLCAAGALAEKVKVCGTSNRLTQGFEDHFLSIQRMYNCEVILGNLEITYVGRN 73
OY 69 ASLSFLDIOEVQGVILAHNQVROVPLQRLRIYRGTOULFEDNVALVLNDGDPMLNTTP 128
DB 74 YDLSFLKTIQEVAGVILALNTVERIPLENQIIRGNALYENTVALAILSN----- 124
OY 129 VTGASPGGLRELOSLRLTEILKGVLIQNPOLCYQDTILWMDI----FKKNQALTLI 184
DB 125 -YGNRTGLRRLPRNLOEILGAVRFNNPLCMMDITIQRDIVQVFMMSMDL--- 180
OY 185 DTNSRACHPCSPMCKGRCWGESSEDCQSLRTVACAGCA-RCKGPLPOYIKANSKEFG 243
DB 181 -QSHPSQCPKCDPSCPNCSWGCGEENCQKLTIKICAGQCSHRGSRGSPDCCHQCAAG 239
OY 244 ITELKHSQCLACLNHNSGICELHCPALVYNTDTFESMPREGYTGASCVTACPNY 303
DB 240 CTGPRESQCLQKQFQDEATKDTCPPLMLYNTPTTYQMDVNEGYSTSGATCVKCKPNY 299
OY 304 LSTVGSCTVLCPLHNOETVAEDGTORCEKSKPCARVCYGLGMBHLREVAVTSANIOE 363
DB 300 VVTDHGSCVRAQGPYYEV-BEDGIRKCKKDDGPRKVCNKGIGIEFPDTSINATNIK 358
OY 364 FAGCKIIFGSLAFPESEFDGDPASNTAPLOEQLQVFETLEITGYLISAMPDLS 423
DB 359 FKCTAISGDHLILPVAFKGSFTPTPLDPRELEILKTVKEITGFLIQAMPDWTDL 418
OY 424 VFQNLQVIRGILNNGASLTLQGLGSLGIRSLRELSGLALIHNTLHLCFVNTVMD 483
DB 419 AFENLEIRGRTHKQGFSLAVGINTISLGRSKESIDSDVILSGRNLCTVANTIMK 478
OY 484 QLFENPQALHTANRPEDECVGEGLACHQLCARGHCGPGPTOCVNCOSFLRGOEYEE 543
DB 479 KLFGPNQKTKIMNRAEKDKAVNVNPNPLCSSGCGPBRDVCQNSRGECEYK 538
OY 544 CRVLQGLREYVNAHCLPCHEPCOPNGSVTCFEPREADQCVACAHYKDPFCVARS 603
DB 539 CNILEGEPRFEVNSECTIOCHPECLPQMMNITTCRGPDNCGICQAHYIDGPHCVTCA 598
OY 604 VKPRLSYPIKPFDEBEACQCPRCPTNCHSCVVDLDDKCC 642
DB 599 IMGENNLT-VMKYADANNVCHLCHANCTYGCAGPGLQGC 636

```

Search completed: July 22, 2003, 09:00:39  
 Job time : 54.3575 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds

(without alignments)  
5267.077 Million cell updates/sec

Title: SEQ4-103-117-14

Perfect score: 6647

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEVIGLDPV 1261

Scoring table: BLOSUM62

Searched: 283224 seqs, 96134422 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: PIR\_73:\*

1: pirl:.\*  
2: pirl:.\*  
3: pirl:.\*  
4: pirl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6708	98.0	1255	1	A24571	protein-tyrosine k
2	5940	86.8	1260	1	TVRTNU	protein-tyrosine k
3	5921.5	86.5	1254	2	I48161	p-185 precursor - f
4	3139	45.8	1210	1	GQHUR	epidermal growth f
5	3108	45.4	1210	2	A53183	epidermal growth f
6	3093.5	45.2	1223	1	TVCHLV	epidermal growth f
7	2986.5	43.6	1308	2	A47253	epidermal growth f
8	2679	39.1	1166	1	S06142	protein-tyrosine k
9	2418.5	35.3	1342	2	A56223	kinase-related tra
10	2334.5	34.1	1339	2	UC4387	epidermal growth f
11	1766.5	25.8	698	1	TVFVLV	protein-tyrosine k
12	1703	24.9	604	1	TVYUH	protein-tyrosine k
13	1655.5	24.2	1330	1	GQFPE	epidermal growth f
14	1647	24.1	544	2	S35745	protein-tyrosine k
15	1640	24.0	545	2	S00727	kinase-related tra
16	1623	23.7	540	2	B44776	protein-tyrosine k
17	1621	23.7	540	1	TVFVB	epidermal growth f
18	1499	21.9	644	1	A36325	epidermal growth f
19	1296	18.9	1333	2	B88257	protein-tyrosine k
20	1296	18.9	1374	2	S70712	protein-tyrosine k
21	1206	17.6	1369	2	S70713	protein-tyrosine k
22	1159	16.9	1717	1	A45558	epidermal growth f
23	1125	16.4	527	2	A42032	epidermal growth f
24	1000.5	14.6	843	2	A27131	epidermal growth f
25	806.5	11.8	346	2	S13807	protein-tyrosine k
26	754.5	11.0	311	2	S13808	protein-tyrosine k
27	727	10.6	1363	2	T43220	insulin-like growth
28	708	10.3	1382	1	INHUR	insulin receptor p
29	701	10.2	1383	2	A36080	insulin receptor p

30	700.5	10.2	1372	2	A34157	insulin receptor p
31	682.5	10.0	1300	2	A36502	insulin receptor-x
32	681	9.9	1607	2	T43212	insulin-like growth
33	671	9.8	1477	2	T18534	protein-tyrosine k
34	669	9.8	1268	2	B36502	insulin receptor-x
35	640	9.3	1367	1	IGHUR1	insulin-like growth
36	629	9.2	1371	2	A33837	insulin-like growth
37	625	9.1	1390	2	T30346	insulin receptor -
38	616	9.0	987	2	A54092	protein-tyrosine k
39	615.5	9.0	2148	1	A56081	insulin receptor -
40	610	8.9	2101	2	S57245	insulin receptor (
41	602.5	8.8	952	2	I50612	protein-tyrosine k
42	597	8.7	987	2	I48652	mouse developmenta
43	594	8.7	984	2	A39753	protein-tyrosine k
44	592	8.6	985	2	I51672	receptor tyrosine
45	591	8.6	987	2	I48953	epn-related recept

## ALIGNMENTS

## RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N.Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein ei

C.Species: Homo sapiens (man)

C.Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C.Accession: A24571, A25491, A44188, B44188, I59509, I57622

R.Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.

A.Title: Similarity of protein encoded by the human C-erb-B-2 gene to epidermal growth

A.Reference number: A24571, MUID:86118663, PMID:3003577

A.Accession: A24571

A.Molecule type: mRNA

A.Residues: 1-1255 <YAM>

A.Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R.Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A.Title: A v-erbB-related protooncogene, C-erbB-2, is distinct from the C-erbB-1/epid

A.Reference number: A25491, MUID:86016729, PMID:2995967

A.Accession: A25491

A.Molecule type: DNA

A.Residues: 737-1031 <SEM>

A.Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R.Cousens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,

Science 230, 1132-1139, 1985

A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chron

A.Reference number: A44188, MUID:86070181, PMID:2999974

A.Accession: A44188

A.Molecule type: DNA

A.Residues: 740-910 <COU>

A.Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A.Accession: B44188

A.Molecule type: mRNA

A.Residues: 1-517, 'RAL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU>

A.Cross-references: GB:M11730; NID:G183986

R.King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A.Reference number: I59509; MUID:85272597; PMID:2992089

A.Accession: I59509

A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 832-909 <REX>

A.Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R.Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcrip

A.Reference number: I57622; MUID:87286898; PMID:3039351

A.Accession: I57622

A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:G183983; PIDN:AA58637.1; PID:G553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERBB2; NGU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 863/3  
 A>Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 kinase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:222-653/Domain: extracellular #status predicted <EXT>  
 F:770-304/Domain: EGF receptor extracellular domain repeat <EB1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <EB2>  
 F:654-675/Domain: transmembrane #status predicted <TM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:68,124,187,259,530,571,629/Binding site: carbohydrate (Aen) (covalent) #status predict  
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:753/Active site: Lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6708; DB 1; Length 1255;  
 Best Local Similarity 98.1%; Pred. No. 1e-267;  
 Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELALCRWGLLALPPGAASVQCTGTDMLRLPASPEHLDMLRHLVYOCQVVGQNL 60  
 DB 1 MELALCRWGLLALPPGAASVQCTGTDMLRLPASPEHLDMLRHLVYOCQVVGQNL 60  
 QY 61 ELTYLPTNASTSLFDIOEVQGYVLIAHNOVRQVPLQRLRYRGVQLFEDNYALAVLDNG 120  
 DB 61 ELTYLPTNASTSLFDIOEVQGYVLIAHNOVRQVPLQRLRYRGVQLFEDNYALAVLDNG 120  
 QY 121 DPLNNFNFTVSWFLRVPKVAS----HLEQLRSLLTELKGGVLLQRPOLCYQDTILMK 176  
 DB 121 DPLNN-----TTVTGASPGGLRELQRLSLTELKGGVLLQRPOLCYQDTILMK 170  
 QY 177 DIFHKNQALATLIDTNRSRACHPCSPMKSGRCWGESSEDDQSLTRTVCAAGCARCKGP 236  
 DB 171 DIFHKNQALATLIDTNRSRACHPCSPMKSGRCWGESSEDDQSLTRTVCAAGCARCKGP 230  
 QY 237 LPTDCHEQCAAGCTGPRHSDCLACLAHFNHSGICELHCPALVYVNTTFESMPNDEGRYT 296  
 DB 231 LPTDCHEQCAAGCTGPRHSDCLACLAHFNHSGICELHCPALVYVNTTFESMPNDEGRYT 290  
 QY 297 FGASCVTACPNYVSTDVGSCTLVGCLPLHNOEYTAEDGTORCKGSKPCARVCGYLGMEHL 356  
 DB 291 FGASCVTACPNYVSTDVGSCTLVGCLPLHNOEYTAEDGTORCKGSKPCARVCGYLGMEHL 350  
 QY 357 REVAVTSANTIOEFAGCKIFGSLAFLPESPDGASNTAPLQEPQLQVFETLEETITYL 416  
 DB 351 REVAVTSANTIOEFAGCKIFGSLAFLPESPDGASNTAPLQEPQLQVFETLEETITYL 410  
 QY 417 YISAMPDLPDLVSQNTQVTRGRILHNGAYSLSLTQIGISWLGRLSRLGSGIALIHH 476  
 DB 411 YISAMPDLPDLVSQNTQVTRGRILHNGAYSLSLTQIGISWLGRLSRLGSGIALIHH 470  
 QY 477 NTHLCFVTVPMDOFLRNPQALHTANRPEDECGEGELACHQICARHCGPGPTQCVN 536  
 DB 471 NTHLCFVTVPMDOFLRNPQALHTANRPEDECGEGELACHQICARHCGPGPTQCVN 530  
 QY 537 CSQFLRQGEVCEECVTLQGLPREVYNARHCLPCHBECOPONGSVTCFPEADQCVACAHY 596  
 DB 531 CSQFLRQGEVCEECVTLQGLPREVYNARHCLPCHBECOPONGSVTCFPEADQCVACAHY 590  
 QY 597 KDPFCVACRCPGCVKPDLSYMPIMKFPBEBGACQPCPINTCHSCVDLDDKCGCPAEORASP 656  
 DB 597 KDPFCVACRCPGCVKPDLSYMPIMKFPBEBGACQPCPINTCHSCVDLDDKCGCPAEORASP 656

DB 591 KDPFCVACRCPGCVKPDLSYMPIMKFPBEBGACQPCPINTCHSCVDLDDKCGCPAEORASP 650  
 QY 657 LSTISAVNGILLVVGIVGVGGILLIKRQOKIRRYTMRRLLOETELVEPLTPSGAMPQA 716  
 DB 651 LSTISAVNGILLVVGIVGVGGILLIKRQOKIRRYTMRRLLOETELVEPLTPSGAMPQA 710  
 QY 717 QMRILKETELARKVYLGSGAFVYKGIWIDGENVKIPVAIKVRENTSPKANKEILDE 776  
 DB 711 QMRILKETELARKVYLGSGAFVYKGIWIDGENVKIPVAIKVRENTSPKANKEILDE 770  
 QY 777 AYVAVGSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVENRGRIGLSODLLMKQOIA 836  
 DB 771 AYVAVGSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVENRGRIGLSODLLMKQOIA 830  
 QY 837 KGMVLEDEVRVYHNDLAARNLVKSPNNVKTITDGLARLLIDFEVYADGKVPIMMA 896  
 DB 831 KGMVLEDEVRVYHNDLAARNLVKSPNNVKTITDGLARLLIDFEVYADGKVPIMMA 890  
 QY 897 LESILRRFTQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTID 956  
 DB 891 LESILRRFTQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTID 950  
 QY 957 VYMTVMKCMIDSECRPFRELVSFSPMARDPQRFVVIQNEIDLGPASPLDSTFYRSILE 1016  
 DB 951 VYMTVMKCMIDSECRPFRELVSFSPMARDPQRFVVIQNEIDLGPASPLDSTFYRSILE 1010  
 QY 1017 DDDMGDLVDAREYLVPOGFFCPCPPAPACGMMHRRHSSSTRGGGLTGLPSEEEA 1076  
 DB 1011 DDDMGDLVDAREYLVPOGFFCPCPPAPACGMMHRRHSSSTRGGGLTGLPSEEEA 1070  
 QY 1077 PRSPPLAPSEAGSDVFPDGLMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYAP 1136  
 DB 1071 PRSPPLAPSEAGSDVFPDGLMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYAP 1130  
 QY 1137 LTCSPOPEYVNPQPVPRPQPSPREGPLPARPAGATTLERAKTILSPGXGVKDVFAFGA 1196  
 DB 1131 LTCSPOPEYVNPQPVPRPQPSPREGPLPARPAGATTLERAKTILSPGXGVKDVFAFGA 1190  
 QY 1197 VENPEYVLTPOGGAPOHPPPAFSPAPDNLYVMODPREBAPSTFGCTPAENPEYL 1256  
 DB 1191 VENPEYVLTPOGGAPOHPPPAFSPAPDNLYVMODPREBAPSTFGCTPAENPEYL 1250  
 QY 1257 LDVPEV 1261  
 DB 1251 LDVPEV 1255

RESULT 2  
 TYRTNU  
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999  
 C:Accession: A24562; A61204  
 R:Barrymann, C.I.; Hung, M.C.; Weinberg, R.A.  
 Nature 319, 226-230, 1986  
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.  
 A:Reference number: A24562; MUID:86118662; PMID:3945311  
 A:Accession: A24562  
 A:Molecule type: mRNA  
 A:Residues: 1-1260 <BAR>  
 A:Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746  
 R:Maui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen  
 Carcinogenesis 12, 1975-1978, 1991  
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals nt  
 2-thiazolyl]formamide or N-methyl-N-nitrosourea.  
 A:Reference number: A61204; MUID:92035293; PMID:1682063  
 A:Accession: A61204  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 637-663, 'V', 665-702 <MAS>  
 A>Note: authors translated the codon GCA for residue 25 as Val  
 C:Genetics:  
 A:Gene: neu



C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP, autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F:1-19/Domains: signal sequence #status predicted <SIG>  
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>  
 F:658-680/Domains: transmembrane #status predicted <TMN>  
 F:723-988/Domains: protein kinase #status predicted <KIN>  
 F:731-733/Region: protein kinase ATP-binding motif  
 F:771,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:693/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:758/Active site: Lys #status predicted  
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.8%; Score 5940; DB 1; Length 1260;  
 Best Local Similarity 87.0%; Pred. No. 2,9e-235;  
 Matches 1098; Conservative 50; Mismatches 108; Indels 6; Gaps 3;

```

Oy 1 MELALCRWGLLALLPGCASTOVCTGDMKRLPASPEHLDMLRHLVYGCQVVGNTL 60
Db 4 MELAMCRWGLLALLPGCASTOVCTGDMKRLPASPEHLDMLRHLVYGCQVVGNTL 63
Oy 61 ELTYLPTNASLFLDIOEVGVYLAHNOVROVLOLRIVRGTOLEFEDNYALAVLDNG 120
Db 64 ELTYVPANASLFLDIOEVGVYLAHNOVRAVLPQRLRIVRGTOLEFEDKVALAVLDNR 123
Oy 121 DPLNNFNNFTVSFMLRVKVASHLEQLRSLTEILKGVLIQRPOLCYQDTILMKDIFH 180
Db 124 DPQDVVAATPG---RTPE--GLRELQRLSLTEILKGVLIQRPOLCYQDVVLMKDVFR 178
Oy 181 KNNQALATLIDTNRSRACHPCSPMKSGSRCWGESSEDCOSLTRTYCAGGACRCKPLPTD 240
Db 179 KNNQALAPVDIDTNRSRACRCPACRCKDNKCGESSEDCQILTGTTCTSCACRCKPLPTD 238
Oy 241 CCHGCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGAS 300
Db 229 CCHGCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGAS 298
Oy 301 CVTACPPNYLSTDVSSCTLVCPNLHNOVTAENGTOCEKCSRPCARVCYGLCMELHREVR 360
Db 299 CVTTCPNYLSTEVSSCTLVCPNNQVTAEDGTORCEKCSRPCARVCYGLCMELHREVR 358
Oy 361 AVTSANIODEFAGCKKIFGSLAFLEPSFDGDPASNTAPLOPELOVPELTETITGLYLSA 420
Db 359 AITSNNOBFDOCKKIFGSLAFLEPSFDGDPSSGIAPLRPELOVPELTETITGLYLSA 418
Oy 421 WPDLSPLDSVFONLQVIRGRILHNGAYSITLQGLGISWLGRLSRLSGLALIHNTYL 480
Db 419 WPDLSPLDSVFONLQVIRGRILHNGAYSITLQGLGISWLGRLSRLSGLALIHNTYL 478
Oy 481 CFVHTVPMDOLEFRNHQALHTANRPDE--CYBEGLAGHQLCARHNCWGPPTQCVCNSQ 539
Db 479 CFVHTVPMDOLEFRNHQALHTANRPDE--CYBEGLAGHQLCARHNCWGPPTQCVCNSQ 538
Oy 540 FLRGOECVBECEGVLCGLPREVYNARHCLPCHPECOPONGSVTCFEPADQCAACHYDP 599
Db 539 FLRGOECVBECEGVLCGLPREVYNARHCLPCHPECOPONGSVTCFEPADQCAACHYDP 598
Oy 599 PCVAVARCSGVKPDLSYMPIMKFPDEBGAQOPCPINCTHSCVDLDDKCGPAEQASPLTS 659
Db 599 SSCVAVARCSGVKPDLSYMPIMKFPDEBGAQOPCPINCTHSCVDLDDKCGPAEQASPLTS 658
Oy 660 IVSAVAVGLLVVVGIVGFGILIKRQOKIRKTKTRLLQETLVEPLTPSGAMPQAOQR 719
Db 659 IIVAVEGVLPLILVVVVGILIKRQOKIRKTKTRLLQETLVEPLTPSGAMPQAOQR 718
Oy 720 ILKETELRKVKVLSGAGFYVKGIMI PDGEVVKIPVAIKVRENTSPKANKEIDEXYV 779
Db 719 ILKETELRKVKVLSGAGFYVKGIMI PDGEVVKIPVAIKVRENTSPKANKEIDEXYV 778
Oy 780 MAGVSPYVSRLLGLCLTSTVOLVTLQMLPYGCLLDHVENRGRLSODLLNMCQIAJGM 839
Db 779 MAGVSPYVSRLLGLCLTSTVOLVTLQMLPYGCLLDHVENRGRLSODLLNMCQIAJGM 838
Oy 840 SYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPIMKMALES 899

```

```

Db 839 SYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPIMKMALES 898
Oy 900 ILRRPETHOSDVMSGVTVWELMTGAKPYDGI PAKEIPDLLEKGERLPQPPCTIDVYM 959
Db 899 ILRRPETHOSDVMSGVTVWELMTGAKPYDGI PAKEIPDLLEKGERLPQPPCTIDVYM 958
Oy 960 IMVKCMIDSECRPFRELVSFSSRMARPPQRFVVIQNEIDLSPASPLDSTFYRSLLEDD 1019
Db 959 IMVKCMIDSECRPFRELVSFSSRMARPPQRFVVIQNEIDLSPASPLDSTFYRSLLEDD 1018
Oy 1020 MCDLYDAEERYLVPOGFFCDDPAPAGGVNHHRRSSSTRSGGDLTGLPSEEEAPRS 1079
Db 1019 MCDLYDAEERYLVPOGFFCDDPAPAGGVNHHRRSSSTRSGGDLTGLPSEEEAPRS 1078
Oy 1080 PLAPSEAGSDVFDLDMGAKGOSLPTHPPSLPQRSEDPPTPLPETGUYAPLVC 1139
Db 1079 PLAPSEAGSDVFDLDMGAKGOSLPTHPPSLPQRSEDPPTPLPETGUYAPLVC 1138
Oy 1140 SPQPEYVNOQDVRPOPSPREBGPLPAARPAGATLSRAKTLSPGKGVNVDVFAFGAVEN 1199
Db 1139 SPQPEYVNOQDVRPOPSPREBGPLPAARPAGATLSRAKTLSPGKGVNVDVFAFGAVEN 1198
Oy 1200 PEYLTPOGGAPOHPHPAPFSAFNLVYMDPPERGAPSPSTFKCTPAENPEYLGUDV 1259
Db 1199 PEYLTPOGGAPOHPHPAPFSAFNLVYMDPPERGAPSPSTFKCTPAENPEYLGUDV 1258
Oy 1260 PV 1261
Db 1259 PV 1260

```

## RESULT 3

148161  
 P-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
 C:Accession: 148161  
 R.Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishika  
 Gene 140, 251-255, 1994  
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Reference number: 148161; MUID:94193007; PMID:7908275  
 A:Accession: 148161  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <RES>  
 A:Cross-references: GB:D16295; NID:9493236; PIDN:BAA03801.1; PID:9747595  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:718-983/Domains: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.5%; Score 5921.5; DB 2; Length 1254;  
 Best Local Similarity 86.7%; Pred. No. 1.6e-235;  
 Matches 1093; Conservative 58; Mismatches 103; Indels 7; Gaps 3;

```

Oy 1 MELALCRWGLLALLPGCASTOVCTGDMKRLPASPEHLDMLRHLVYGCQVVGNTL 60
Db 1 MELAMCRWGLLALLPGCASTOVCTGDMKRLPASPEHLDMLRHLVYGCQVVGNTL 60
Oy 61 ELTYLPTNASLFLDIOEVGVYLAHNOVROVLOLRIVRGTOLEFEDNYALAVLDNG 120
Db 61 ELTYVPANASLFLDIOEVGVYLAHNOVRAVLPQRLRIVRGTOLEFEDKVALAVLDNR 120
Oy 121 DPLNNFNNFTVSFMLRVKVASHLEQLRSLTEILKGVLIQRPOLCYQDTILMKDIFH 180
Db 121 DPLNNFNNFTVSFMLRVKVASHLEQLRSLTEILKGVLIQRPOLCYQDTILMKDIFH 174
Oy 181 KNNQALATLIDTNRSRACHPCSPMKSGSRCWGESSEDCOSLTRTYCAGGACRCKPLPTD 240
Db 175 KNNQALAPVDIDTNRSRACRCPACRCKDNKCGASPEDCQILTGTTIAPRAVPAARARLPTD 234

```

```

QY 241 CCEOCACGCTGPKHSDCLACLFPHNSGICELHCPALVTYNTDTFESMPNREGRYFGAS 300
DB 235 CCEOCACGCTGPKHSDCLACLFPHNSGICELHCPALVTYNTDTFESMPNREGRYFGAS 294
QY 301 CTTACPYNYLSTDVGSCTIWCPLHNOEVTAEADTQCEKSKCPARCYGALGMEHREVR 360
DB 295 CTTCPYNYLSTEVGSCTIWCPLHNOEVTAEADTQCEKSKCPARCYGALGMEHREVR 354
QY 361 AATSAIOEFACGCKTIFGSLAFLPESPDGPANTAPLOEOLOVEETLEITGTYLISA 420
DB 355 AATSAIOEFACGCKTIFGSLAFLPESPDGPANTAPLOEOLOVEETLEITGTYLISA 414
QY 421 WPDLSPLDSVFOQLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELSGSLALIHNTHL 480
DB 415 WPDLSPLDSVFOQLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELSGSLALIHNTHL 474
QY 481 CFVHTVPMDOQLFRNPHQALHTANRDEDECVGEGLAGCHQICARGHCGMPGPTQCVNCSOP 540
DB 475 CFVHTVPMDOQLFRNPHQALHTANRDEDECVGEGLAGCHQICARGHCGMPGPTQCVNCSOP 534
QY 541 LAGQECVEECRVLOGLPREVYNARHCLPCHPECOQNGSVTCRGPENADOCVACAHYKDDP 600
DB 535 LAGQECVEECRVLOGLPREVYNARHCLPCHPECOQNGSVTCRGPENADOCVACAHYKDDP 594
QY 601 FCVARCPGKVPDLSTYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCPAEOGASPLTST 660
DB 595 FCVARCPGKVPDLSTYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCPAEOGASPLTST 654
QY 661 VSAVVGILLVVLGVVFGILIKRROOKIRKRYTMRLLQETELVEPLTPSGAMPNOQMRI 720
DB 655 IATVVGILLVVLGVVFGILIKRROOKIRKRYTMRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKXKV/GSGAGCTYTKYKIWDGNNVKIPVAIKYLRNTSPKANKELLDAYVM 780
DB 715 LKETELRKXKV/GSGAGCTYTKYKIWDGNNVKIPVAIKYLRNTSPKANKELLDAYVM 774
QY 781 AGVGSFYVRLTGLICTSTVQVLTQMLPYCCLDHYRENNRGSGODLNMCKQIAKMS 840
DB 775 AGVGSFYVRLTGLICTSTVQVLTQMLPYCCLDHYRENNRGSGODLNMCKQIAKMS 834
QY 841 YLEDVFLVHRDLAARVLYKSPNHVKITDFGLARLLDIDETEHADGKVPIMWALEST 900
DB 835 YLEDVFLVHRDLAARVLYKSPNHVKITDFGLARLLDIDETEHADGKVPIMWALEST 894
QY 901 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIAREIPDLLEKGERLPQPPICITIVYMI 960
DB 895 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIAREIPDLLEKGERLPQPPICITIVYMI 954
QY 961 MVKCMNIDSECRPRFELVSEFSRMAADPQRFVVIQNEDELGPASPLDSTFYRSLDEDDM 1020
DB 955 MVKCMNIDSECRPRFELVSEFSRMAADPQRFVVIQNEDELGPASPLDSTFYRSLDEDDM 1014
QY 1021 GDVDAEELVLPQGGFCPPDPAGQGVHHRSSSTRSGGDLTLGLEPSEEARSP 1080
DB 1015 GDVDAEELVLPQGGFCPPDPAGQGVHHRSSSTRSGGDLTLGLEPSEEARSP 1074
QY 1081 LAPSEAGSDVPDGLGMAKAGLOSLPTHDPSPLORYSDPVPVLPSEEDGVAPLTC 1140
DB 1075 LAPSEAGSDVPDGLGMAKAGLOSLPTHDPSPLORYSDPVPVLPSEEDGVAPLTC 1134
QY 1141 POPEVYNQDVPQPSPREGRLPAARPAATLERAKTUSPGNGVGVKDFAFGAVENP 1200
DB 1135 POPEVYNQDVPQPSPREGRLPAARPAATLERAKTUSPGNGVGVKDFAFGAVENP 1194
QY 1201 EYLTPOGGAAPQHPAPASPADNLYYMDQDPEPGAPSTKGPPTAENPYLGLDVP 1260
DB 1195 EYLTPOGGAAPQHPAPASPADNLYYMDQDPEPGAPSTKGPPTAENPYLGLDVP 1253
QY 1261 V 1261
DB 1254 V 1254

```

## RESULT 4

## GOME

epidermal growth factor receptor precursor - human

N/Containing: protein-tyrosine kinase (EC 2.7.1.112) erbB

C/species: Homo sapiens (man)

C/date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999

C/accession: A00641; A25772; S30024; A3672; A0642; A3615; A23062; A05281; A60143;

R/Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.

rg, P.H.

A:feature 309, 418-425, 1984

A:title: Human epidermal growth factor receptor cDNA sequence and aberrant expression

A:reference number: A00641; MUID:84219729; PMID:6328312

A:accession: A00641

A:molecule type: mRNA

A:residues: 1-1210 &lt;UL&gt;

A:cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g57924

A:note: the authors translated the codon AAG for residue 540 as Asn

R/Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.

Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985

A:title: Characterization and sequence of the promoter region of the human epidermal

A:reference number: A25772; MUID:85270438; PMID:2991899

A:accession: A25772

A:status: translation not shown

A:molecule type: DNA

A:residues: 1-29 &lt;ISH&gt;

A:cross-references: GB:M1234; NID:g181981; PIDN:AAA52370.1; PID:g553272

R/Haley, U.; Whiteley, N.; Bennett, P.; Kitchington, D.; Ulrich, A.; Waterfield, M.

Oncogene Res. 1, 375-396, 1987

A:title: The human EGF receptor gene: structure of the 110 kb locus and identification

A:reference number: S30024; MUID:88217333; PMID:3329716

A:accession: S30024

A:molecule type: DNA

A:residues: 1-29 &lt;HA2&gt;

A:cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119

R/Haley, J.D.; Waterfield, M.D.

J. Biol. Chem. 266, 1746-1753, 1991

A:title: Contributory effects of de Novo transcription and premature transcript termi

A:reference number: A38672; MUID:91107677; PMID:1988448

A:accession: A38672

A:molecule type: DNA

A:residues: 1-29 &lt;HML&gt;

A:cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271

R/Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.

Nature 309, 806-810, 1984

A:title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN

A:reference number: A00642; MUID:84245835; PMID:6330563

A:accession: A00642

A:molecule type: mRNA

A:residues: 'RCAMRA', 150-187, 'KSVIOAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-

, '98-99, 'TD', 802-811, 'R', 813-942 &lt;XUY&gt;

A:experimental source: A431 human carcinoma cells, which have large numbers of EGF re

S/lin, C.R.; Chen, W.S.; Knutger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,

Science 224, 843-848, 1984

A:title: Expression cloning of human EGF receptor complementary DNA: gene amplificati

A:reference number: A43615; MUID:84196372; PMID:6326261

A:accession: A43615

A:molecule type: mRNA

A:residues: 713-964 &lt;LIN&gt;

A:experimental source: epidermoid carcinoma cell line A431

R/Stimmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.

Biochem. Biophys. Res. Commun. 124, 125-132, 1984

A:reference number: A23062; MUID:85046483; PMID:6093780

A:accession: A23062

A:molecule type: mRNA

A:residues: 1028-1210 &lt;SIM&gt;

R/Weber, W.; Gill, G.N.; Speiser, J.

Science 224, 294-297, 1984

A:reference number: A05281; MUID:84172183; PMID:6324343

A:accession: A05281

A:molecule type: protein

A:residues: 25-30, 'S', 32-51, 454-467 &lt;WEB&gt;

R; Russo, N.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.  
 J. Biol. Chem. 260, 5205-5208, 1985  
 A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor  
 A:Reference number: A60143; MUID:85182850; PMID:2985580  
 A:Accession: A60143  
 A:Molecule type: protein  
 A:Residues: 740-744, 'X', 746-747 <RUS>  
 R; Mroczkowski, B.; Mosig, G.; Cohen, S.  
 Nature 309, 270-273, 1984  
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superec  
 A:Reference number: A38023; MUID:8491554; PMID:6325948  
 A:Contents: annotation; receptor activity  
 A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
 R; Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
 Cell 59, 33-43, 1989  
 A:Title: Functional independence of the epidermal growth factor receptor from a domain  
 A:Reference number: A33311; MUID:90003233; PMID:2790960  
 A:Contents: annotation; internalization signal  
 A:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor  
 C:Genetics:  
 A:Gene: GDB:EGFR  
 A:Cross-references: GDB:120610; OMIM:115550  
 A:Map position: 7p12.3-7p12.1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-120/Product: EGF receptor #status predicted <KMT>  
 F:25-645/Domain: extracellular #status predicted <EXT>  
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>  
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:668-668/Domain: transmembrane #status predicted <TM>  
 F:669-1210/Domain: intracellular #status predicted <INT>  
 F:710-975/Domain: protein kinase homology <KIN>  
 F:718-726/Region: protein kinase ATP-binding motif  
 F:999-1046/Region: coated-pit mediated internalization signal  
 F:1047-1210/Region: inhibitory  
 F:148,175,352,413,444,528,603/Binding site: carbohydrate (asn) (covalent) #status predic  
 F:745/Active site: Lys #status experimental

Query Match 45.8%; Score 3139; DB 1; Length 1210;  
 Best Local Similarity 49.3%; Pred. No. 1.5e-121;  
 Matches 627; Conservative 179; Mismatches 351; Indels 116; Gaps 22;

11 LLLALLPFGAA--STOVCTGDMKRLRPASPEYTHLDMRLHYOGQVQVQNELTYLPTN 68  
 14 LLAALCPASRALEEKKVCQGTSTNKLQTGTREDHFLSLQRMNNEVVLGNLEIYYQGN 73  
 69 ASLSFLDIOGVQGVVLIANQVROVPLORLRIYVGTQLFEDNVLAVALVNDGPIANFN 128  
 74 YDLSFLKTIQVAVGVLIANTVERIPLENLIIRGNMYENSVALAVLSYND----- 126  
 129 FTVSFVLRVPRVKSASHLEOL--RSITELIKGGVLIQRMPOLCYOTIIMKDI FHKNO 186  
 127 -----ANKTGLKELPMRNLOETILHGAIVRSPNPAICNWSIOWRIDVSSDPLSN 175  
 187 LTLIDTRSRACHPCSPWCKSGRCWGESSEDCOSILRTVYAGGCA--RCKGPLPTCCHQ 245  
 176 MSMPFOHNLGSCQKCDPCSPGSCGAGEENCQKLTILICAQCSGRKGRKSPSCCINO 235  
 246 CAAGCTPRKXBDCLACHFNHSGICELHCPALVYNTDTFESMPNBEGRYTFGASCVTAC 305  
 236 CAAGCTGPRESDCLVCRKFRDEATCKDTCPRMLYNTTYOMDVAVPEKYSFGATCVKCC 295  
 306 PYNLTSTDVGSCTLVCPLEHNOEVTAEADGTORCEKSKCAVYCVGLGMEHLREVAATVSA 365  
 296 PRNVVTDHSGCVAAACADSTYEM--BEDGVKCKCKCEGCRKVCNGIGIGEFKDSLSINAT 354  
 366 NIOEFACCKIKFGSLAFLESFDDDPASNTAPLOPEQLOVEETLEITGVLYISAMPDL 425  
 355 NIKFKKCTKSISGDLHILPVAFRGDSFTHTPRLDQEDILKTYKEITGFLIQAMPNR 414  
 426 PDLVSPNLOVIRIRILHNGAYSLTLQGLISWLGRLSLRGLGSLALIHNTLCTVHT 485

415 TDLHAFENLEIRGRTKQHGQFSLAVSINITSLSRLSKEISDGDVILISGNRLCYANT 474  
 486 VPMDDLFRRPHQALHTARREDCEGBELACHQICARHNCRGRTQVCNCOFLRGOE 545  
 475 INMKCLFGTSGQTKTKISNRGENSCATQVCHALCSPEGCWPERDCVSCNVSRRGE 534  
 546 CVEBCRVLOGLPREVYNARHCLPRHPECOPONGSVTCFPEADQVCACAHYKPPCVAR 605  
 535 CYDKCKLLEGEPRFVENSECICQHPECIPROMNITTCGRDNCICQAHYIDGPHCVKT 584  
 606 CPGVKRPDLSTYMPIMKFPDEBACQPCPINCISCVLDLDCGCPABORASPLTSISAVV 665  
 595 CPAGVWGEVNTL-VKXVADAGVCHLCHPNCYCTGCTGPELBCPTGPKIP--SIATGV 651  
 666 G---LLVYVGVGVGILIKRQOKIRKTKTMRLLQETLVEPLTSSGAMPNOQRILK 722  
 652 GALLLLVVALIG---LEMRHRHIVRKTLRRLLOERLVBPLTPSGAPNOLRLIK 708  
 723 ETELKRVKVLGSGAGTVYKGIPIPDGENVKIPVAIKVRENTSPKANEILDEAYVMG 782  
 709 ETEFKKIKVLGSGAGTVYKGIPIPDGENVKIPVAIKELRENTSPKANEILDEAYVMG 768  
 783 VGSPPVSRLLGICLTSTVQLVTQMLPYGCLLDHVENRGLSGODLNNCMQIAKMSYL 842  
 769 VDNPHVCRLGICLTSTVQLVTQMLPYGCLLDHVENRGLSGODLNNCMQIAKMSYL 828  
 843 EDVRLVHBDLARNLVVSPHNVKITTDEGLALLDIDETRYADGSKVPIKMALESILR 902  
 829 EDRLVHBDLARNLVVSPHNVKITTDEGLALLDIDETRYADGSKVPIKMALESILR 888  
 903 RRTFQSVWVSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPICTIDVYMINV 962  
 889 RLYTQSDVWVSGVTVWELMTFGSKPYDGIIPAREIPDLLEKGERLPORPICTIDVYMINV 948  
 963 KCMWIDSECRPRFRELVEFSRMAEDPORFVVIQ--NEDLGRASPDLSTFYRSLDEDDMG 1021  
 949 KCMWIDSECRPRFRELVEFSRMAEDPORFVVIQ--NEDLGRASPDLSTFYRSLDEDDMG 1008  
 1022 DLVDAEETLVPOQGFCCDDPAGAGMNNHRRSSSTSGGDDLTLGLEPSEEARSP 1081  
 1009 DVVDAEETLVPOQGF-----SSPSTSRPL 1034  
 1082 APSEGAGSDVPDGDGCMGAAGKLOSLPTHPSPRLORYSDDPTVPLPSR--DGVVAPRLC 1139  
 1035 LSSLATSN--NSTYACLDRLNGLOSCPIKEDSFLRYSDDPTGALTEDSIDTFL----- 1087  
 1140 SPOPEYVQPDVPRPSPRREGRLPARAPAGATLERAKTSLSPGKGVVVDVFAFGAVEN 1199  
 1088 -PVEYINQ-SVPRKPAQSVQRPVYHNPRLN-----ASRDPHYD--PHSTAVGN 1135  
 1200 PEYL-TPGGAAPQHPRRPAPSPADNLTYYDQ-----DP-----DERGAPST 1242  
 1136 PEYLVNTVO-----PTCVNSTPDSPAHMAQKSHQISLDNPYQODFPKPAKPNGI 1186  
 1243 FKGTPTAENPEYL 1255  
 1187 FKGS-TAENAYEL 1198

RESULT 5  
 A53183  
 C:Species: Mus musculus (house mouse)  
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999  
 C:Accession: A53183; A43818; S24942; A28941; S45325; I49643  
 R:Unetleke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Barp, H.S.; Jenkins, N.A.  
 Genes Dev. 8, 399-413, 1994  
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor  
 A:Reference number: A53183; MUID:94170986; PMID:8125235  
 A:Accession: A53183  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <LUE>  
 A:Cross-references: GB:U03425

R.A.Vivi, A.; Lax I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B. Oncogene 6, 673-676, 1991

A>Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site

A:Reference number: A43818, MUID:91232866; PMID:2030916

A:Accession: A43818

A:Molecule type: mRNA

A:Residues: 1-714 <AVI>

A:Cross-references: GB:X59698

R:Eisinger, D.P.; Serrero, G

submitted to the EMBL Data Library, June 1992

A:Reference number: S24942

A:Accession: S24942

A:Molecule type: mRNA

A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>

A:Cross-references: EMBL:212608

R:Heisermann, G.J.; Gill, S.K.

J. Biol. Chem. 263, 13152-13158, 1988

A>Title: Epidermal growth factor receptor threonine and serine residues phosphorylated

A:Reference number: A28941; MUID:88330814; PMID:3138233

A:Accession: A28941

A:Molecule type: Protein

A:Residues: 689-694, 'K', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009, R.Hilbs, M.L.; Dunn, A.R.; Alexander, W.S.

submitted to the EMBL Data Library, April 1994

A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor

A:Reference number: S45325

A:Accession: S45325

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-971, 'K', 973-1210 <VER>

A:Cross-references: EMBL:X78887; NID:G488830; PIDN:CAA55587.1; PID:G488831

R:Patla, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.

Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993

A>Title: Expression of the epidermal growth factor receptor gene is regulated in mouse

A:Reference number: 149643; MUID:93126380; PMID:7678348

A:Accession: 149643

A:Status: translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 12-20, 22-132 <RES>

A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201

A:Genetics:

A:Gene: EGFR

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphotrc

F:1-24/Domain: signal sequence #status predicted <SIG>

F:648-670/Domain: transmembrane #status predicted <TM>

F:712-977/Domain: protein kinase homology <KIN>

F:720-728/Region: protein kinase ATP-binding motif

F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental

F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental

F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental

F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental

F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3108; DB 2; Length 1210; Best Local Similarity 49.1%; Pred. No. 2.7e-120; Matches 629; Conservative 171; Mismatches 360; Indels 120; Gaps 24;

Qy 11 LLLALLPFGAA--STVCTGTDMLRLPASPETHLDMRLHYOCGVVOCNLELTYPTN 68

Db 14 LTLTALCAAGGAALEEKVKVCGGTSNRLTQLGTFEDHFLSLQRYMNCCEVVLGNLEITTYQRN 73

Qy 69 ASLSFLDIOGVGVVLIHNVQVQPLQRLRIYRGQLPFDNVALAVLNDGDIANNFN 128

Db 74 YDSFLKTIQDSVAGVLIALTAVTERIPLENQIIRGNALYENTYALAILSYG----- 126

Qy 129 FTVSFWLVRPKVASHTLEQL--RSLTETLKQGVLIQRNPOLCYODTILMKDI-----FKKN 182

Db 127 -----TNRGTRELPMRNLIQELILGAVFNSNNPILCNMDITIQMDIYQNVFMSN 175

Qy 183 NQALTLIDNTRSAACHRCFMCCKGSRCKWSSSEDCQSLRTTVACGGA-RCKGPLPTDC 241

Db 176 NSMPL----QSHFSSCPKCDSPSCNGSGCWGGEGNCKLTKTITCAQDCSHRCRSPSDC 231

QY	242	CHBCCAGCTBPKNKSDCLACHFNHSGI	CELYCHALYVYNNDTFESMNP	PERGYFGAC	301		
QY	242						
Db	232	CHNOACAGCTPRBSDDLVCOXKFODEATC	KOTCPRLMYNPTTIOMDVNPBCKS	FGATC	291		
QY	302	VTACRPYVNLSTDVSSCTLVCPHNOE	VYAEADGTORCEKSKP	PCARVCYGLGMEHLYREVA	361		
QY	292						
Db	292	VKCKPRNVVYVTHDSCSVACGPRDYEV	-EDGQIRKCKKCDQPCRKVNCIGI	IGERKDTLS	350		
QY	362	VTSANIOEFACCKKI	FGSLAF	PESFGDPSAMNTAPLOPOLQVF	LEETITGYLYISAM	421	
Db	351	INANIKIKFKYCTAISGDHLIP	PAFKDSSFTTRPPRLDPRLEILIKY	KEITGFFLLIDAM	410		
QY	422	PDLSLDLSVFNOLVYIRGRILHNA	AVSLTLOGIGISWGLSLEISGLAL	IHNHTJLC	481		
Db	411	PDNMTDHAFFNLEIIRKTYKOHQ	OFSLAVVGLNITSIGLSLKEI	SOGDYIISGNRYLC	470		
QY	482	FVHTVPMDOLEFRNPHOALLTANP	REDECSVEGLACHOLCAR	GHCWGMPPOCVNCSOFL	541		
Db	471	YANTINMKKULFGTNGOKTKIMNN	BAEKDCKAVNHVCNPLCS	EGCKMGEPRDVCSCVWS	530		
QY	542	RGQECVECEVLOGLPREYVNAH	RCPCHECQORNGSVTCS	FGREADOCVACAHYKDPF	601		
Db	531	RGRECSVECKNILEEBREBEFVENSE	CIQCHPECLPRAMNITCTG	SRGPDNCCI	OCANHYIDSPH	590	
QY	602	CVACPSGVKRDLSYMPRIKFPDE	GCACQRPINTHSCVDL	DXDGCBAEQASLTSIV	661		
Db	591	CVKTCSPAGIMENNTL-VMKYAD	ANNVCHLCHNACTYGCAC	PRQGC	CVWSPSPKPI	PS1A	649
QY	662	SAVGIILLVNLGVVFGI-LIKR	QOQKIRKTYMRLLQETELVE	LTGSGAMPNOAORI	720		
Db	650	TGIYGLLFIYV-VALGIGL	PMKRRIYKRTKLRLDRE	LYVELTPTSGEAPNOAH	LRI	708	
QY	721	LKETELRKVULVSGAGCTVYKGI	WIPDGENVKI	PVAIKVLR	RENTSPRANKELIDEAYVM	780	
Db	709	LKETEFKKIKYLGSGAGFCTVYK	GLMIPEGEKVI	PVAIKELREATS	PRANKELIDEAYVM	768	
QY	781	AGVSPYVSLTIGCTSTVOLVQ	LMYPCGLLDHYENRGR	LSQDILNNCMQIAKMS	840		
Db	769	ASVNPVHCRLGLICTSTVOLIT	QMLMPYGCLLDYVREH	KONISQYULNNCVOIAKMN	828		
QY	841	YLEDVRLVHRLAARNVULVKS	PNVVKITDFGLARL	LDIDETEHADGSKVPI	KMMALESI	900	
Db	829	YLEDBRLVHRLAARNVULVKT	POHVKITDFGLAKLGA	BEKYEYHAEGBKVI	KMMALESI	888	
QY	901	LRRBFTHQSDVMSYGVTV	WELMTFGAKPYDGI	PARREI	PDLLEKGBRLPQRPITIDVYMI	960	
Db	889	LHRITYHQSDVMSYGVTV	WELMTFGSKPYDGI	PASDIS	ISILEKGBRLPQRPICITIDVYMI	948	
QY	961	MVKCMIMIDSECRPRFREL	VEFSFMSARDPOGFVYIQ	-NEDJG	SPASPLDSTYRSILEDD	1015	
Db	949	MVKCMIMIDADSRPRFREL	IEFSQMARPOFVYLIQ	GERMNI	LPSPDSNFYRALIMDEED	1008	
QY	1020	MGDVLDAEYLVPOOGFFCPD	PARAGAGVMHHRSSSTR	SGGDLTGL	LEPSEBEARS	1079	
Db	1009	MEDVYVDAEYLIPOOGFF-----	-----NSRST-----	-----SRT	-----	1033	
QY	1080	PLABSEGAGSDVFDGLG	MGAKGLQSLPTHDP	SPRLQYSEDPVLP	LPSET--DGYVAPL	1137	
Db	1035	PLLSLSLATSIN-----	NSTVACIN	NGSCRYKEDAF	LQRYSSDPGALVEDNIDAF	1087	
QY	1138	TCSQOPREYVNO	PDVPRQSPREBGLPAAR	AGATLE	RAKTLSTGKQGVKVDVAF	GGAV	1197
Db	1088	---SVPEYVNO-SVPKBP	AGSVQNPVHNQPLH---	-----	AGRDLAHYON--PHSNV	1133	
QY	1198	ENPEYLT-TPQGA	PARPHRPA	SPAFDNLYVMO-----	DP-----	PERGAPR	1246
Db	1134	GNPEYLYTAAQ-----	PTCLSSG	GRNSPALMIQKSHQ	MSLNDPVOQDFF	FKETKRN	1184
QY	1241	STFKGTP	TAENPEYLTGLDVP	1260			
Db	1185	GIFKG-PTAE	VAEYTLRAVP	1203			

RESULT 6

TYCHLV

epidermal growth factor receptor precursor - chicken

N/Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C/Species: Gallus gallus (chicken)

C/Date: 28-Feb-1986 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000

C/Accession: A27720; A00643

Relax. I.: Johnson, A.; Hawk, R.; Sap, J.; Belloc, F.; Winkler, M.; Ullrich, A.; Vennstr

Mol. Cell. Biol. 8, 1970-1978, 1988

A/Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A/Reference number: A27720; MUID:88261272; PMID:3260329

A/Accession: A27720

A/Molecule type: mRNA

A/Residues: 1-1223 <LAX>

A/Cross-references: GB:M20386

R/NILsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M

Cell 41, 719-726, 1985

A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and proc

A/Reference number: A00643; MUID:95228222; PMID:2988784

A/Accession: A00643

A/Molecule type: mRNA

A/Residues: 585-1223 <NTL>

A/Cross-references: GB:M10066

C/Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor

specific protein kinase

F/1-30/Domain: signal sequence #status predicted <SIG>

F/31-1223/Product: epidermal growth factor receptor #status predicted <MAT>

F/31-654/Domain: extracellular #status predicted <EXT>

F/81-307/Domain: EGF receptor extracellular domain repeat <EE1>

F/197-610/Domain: EGF receptor extracellular domain repeat <EE2>

F/655-677/Domain: transmembrane #status predicted <TM>

F/718-1223/Domain: intracellular #status predicted <INT>

F/719-984/Domain: protein kinase homology <KIN>

F/727-735/Region: protein kinase ATP-binding motif

F/136-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #

F/197,650/Binding site: carbohydrate (Ser) (covalent) #status predicted

F/687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F/754/Active site: Lys #status predicted

F/1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.24; Score 3093.5; DB 1; Length 1223;

Best Local Similarity 48.04; Pred. No. 1,3e-119;

Matches 626; Conservative 175; Mismatches 351; Indels 151; Gaps 24;

8 RWGLLALLPPGAA-----STOVCTGDMKRLRPASPEHILMDLRHLTYOGCGVVGNE 61

13 RGAVALVLLLLGLGVALCSAVEKKVCGGTNNKLTQGHVEDHFTSLQRMVNNCEVALSNLE 72

62 LTYLPNNASLSFLQDIOEVGGYVLIAHQVQRPQLQRLRYRGTOLFEDNVALAVIDNGD 121

73 IYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENIQIIRGNVLYDNSFALVALSN-- 130

122 PLANNFNFTVSPMLRVPKVASHLEOLRSLTEILKGGVLIOENPOLCYQDTLTKMOIFPK 181

131 -----YHNKTKQGLRELPMKRLISEILNGVKYKISNPKLCNMVDVLMNDIIT 177

182 NNQLALFLID-TNRSRACHPCSPMKCSRCMGESSEDCQSLRTVYACGCA-RCKGPLPT 239

178 SRK-PLTVLDFASNLSSPCPKHPCTEDHCGAGEBQNCOTLTKVICAQCCSRCKGVPS 236

240 DCCHCCAGCGTGRKXSDCLALHFNHSGICELCPALVTYNTDTFESHPNREGRTFGA 299

237 DCHNQCAAGCTGPRSDCLACKRFDPATKDTCPPLVLNPTTYQMDVNBEGKXSFSA 296

300 SCVTACPYNYLSTDVSGCTLVCPALHNOEVTVAEDGTQRECKSKPCARVCYAGMEHLREY 359

297 TCVRECPHYVTVTHDSCVASKNDTYEV-ENGVRRKKKKDGLCSKYCNIGIGIGELKEI 355

360 RAVTSANIDEPACCKKIFGSLAFLPSPFDGDPASNTADLPQBOLOVFETLEETGYLYIS 419

Db 336 LSIATNATIDSRKCTKINGDVSILPVAFLGAPFKTLPLDPKLLDVFRTVVEISGFLIIQ 415

Qy 420 AMPDPLDPLSVFONLOVIRGRILHNAGVSLTLOGLISWLGSLRSIRELSGLALIHNNH 479

Db 416 AMPDNPATDVAFENLEIIRGRTHQHGQSLAVLVNKLQISGLRSGLKEISDGDIAIMKKN 475

Qy 480 LCFNHTVWDOLFRRPHQALLHTANRREDCEVSGGLACHQLCARGHCMPGPTCVNCSQ 539

Db 476 LCVADTNMNRSLFATOSQKTKIIONRMKNDCTADRHVCDPLCSVVGCGMPFPHCFSCRF 535

Qy 540 FLRGECEVEERVLQGLPREYVNAHNLPCRPEDOPONG---SYTCGPRADOCVACHY 596

Db 536 FSROKECKQCNIIQGEFRERFERDSKCLPCSECLVQNSTAYNTTCGPGPHCMKCAHF 595

Qy 597 KDPEPCVACRCSGVKPDLSYMPIMKFPDEEGACPCPINTCTSCVDLDDKDCAPGORASP 656

Db 596 IDPHPCVACACAGVLGENDTL-VKRYADANAVCQLCHPNCTRGCKGPLEBCEP---NSK 651

Qy 657 LTSIVSAVY-GILLVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQ 715

Db 652 TPSIAGVAGGLCLLVVGLIGLGLYRRR-HIVKRTLRLLORELEVEPLTPSGEAPNQ 710

Qy 716 AQMLIKETELARKYVVGSGAFGVYKGIWIPDEBNKIPVAIVLEBNSPKANKELD 775

Db 711 AMLRLKETEKRYKVVVGSGAFGVYKGLWIPDEBKYPVAILEBNSPKANKELD 770

Qy 776 EAYVWAGVGSFYVSRLLGI CLTSIVQVLTQVLTQAMPGLCLLDHRENRGLSGODLLNMCQI 835

Db 771 EAYWASVNDNHVCRLLGLICLTSIVQVLTQVLTQAMPGLCLLDYREHKNDIGSOYLNMVQI 830

Qy 836 AKGMSYLEDELVRHRLDAARNLVKSPNHVKITDFGLARLLIDETEVHADGKVP1KMM 895

Db 831 AKGMVYLEERLVRHRLDAARNLVKTPQHVKITDFGLAKLLGADKEVYHAEQKVP1KMM 890

Qy 896 ALESILRRRFFHNSDWSYGVTVWELMTFGAKKPRDGI PAEELPRLKEGRLPQPICTI 955

Db 891 ALESILHRITYHNSDWSYGVTVWELMTFGSKPRDGI PAEELSSVLEKGERLPQPICTI 950

Qy 956 DVYIMVYKMWIDSECRPRPRELSEFSRMAARDQRFVVIQ-NEDLCPASPLDFTFYSL 1014

Db 951 DVYIMVYKMWIDSECRPRPRELSEFSRMAARDQRFVVIQ-NEDLCPASPLDFTFYSL 1010

Qy 1015 LEDDDMGDLVDAEYLVLPQGFPCRPDPAFGAGVNHNRHSSSTRSGGSDLTGLBPSEE 1074

Db 1011 MEEDMEDIVDAEYLVLPQGF-----NSPST----- 1038

Qy 1075 EAPRSPL-----APSEAGSDVDFGDLGMAKGLQSLPHNDSPLORYEDPTVPLPSE 1129

Db 1039 --SRTPLSLLSATSNNSATNCID-----RNGQCHPRVEDSFVORYSDSPGFLLEE 1088

Qy 1130 T--GGVAVRLPSCQRPQEVNQDVRQPRSPRESERLAARPAAGTLERATLSPGCKGV 1187

Db 1089 SIDSGFL-----PAPEVNO--LMPKKPS-----TAAVNOQIY 1119

Qy 1188 KDVF-----AFGAIVENPEYLTPOGGAAROPHPRAPSPAFMDLYVWDQ- 1231

Db 1120 NNISLTAISKLPMSRKYONSHSTAVNDPRL-----NTNQSPLAKTVFESSPTVIQS 1171

Qy 1232 -----DPE-----KQAPSTFGTGTPTAENPEYLGIDVP 1260

Db 1172 GNMQIINDNPYQODFLPNETKPMGLKLVPALENPEYLRVAAP 1214

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

CSpecies: Homo sapiens (man)

CDate: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999

CAccession: A47253

Proteinman, G.D.; Culousoou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal



Matches: 575; Conservative 163; Mismatches 387; Indels 152; Gaps 29;	
Qy	4 AALCRWGLLALIPGAASL---QVCTGDMKRLRPASPEHLDMLRHLVYCCGVYQGN 59
Db	8 AALLQ--LLLVLSIRSCSTDPDRKVCQGTSMQMT--LDNHYLKKMKXVSSGCVNLEN 62
Qy	60 LETLITPTNASTSLFDIOIEVQGVYLIANQVROVLPQRLRIYRGQLPEDNYALAVLDN 119
Db	63 LEITTOENODLSFLDSIQEVGVYLIANNEVSTIPLVNLRLIRGNLYEGNTLLVMSR 122
Qy	120 --GDPLENNFNFTVSEFMLRVPKVASHLBOLR--SLTEILKGVLIQBNPOLCYODTIIM 175
Db	123 YQKNSS-----PDVYQVGLQLOLSNLTLSLGGVKKVSHNPLCNVETIM 169
Qy	176 KDI FHKNNQALTLIDTNRSRACHPCSPMCKSGSRCKGSESDCQSLTRTVACGC-ARCK 234
Db	170 WDIVKTSNPTMMLLFPAFEEOQCKDHCCVNGSCVAPRGHCQKFTKLLCAQCNRRCR 229
Qy	225 GPLPTDCCHEGCAAGCTGPKHSDCLALPHNSGIELLCPLAVTYNTDTFESMPDEGR 294
Db	230 GKPLTDCCNEHCAGCTGPRATDCLACRDFNDGTCOTCPPKIYDIYSHQVNDPNIK 289
Qy	295 YTFGASCVTACPYNYLSTDVSGCTLVCPPLHNOEVTAEQGTORCEKCSKPCARCYGLME 354
Db	290 YTFGAACVKECSNIVYTE-GACVRSCKSGMLEVD-ENGRKSKPCDCGCPKVCIGIG 347
Qy	335 HLREVRATVSANIQEFAGCKKIFGSLAFPESEFDGPASNTAPLOEOLQVETLEITG 414
Db	348 SLSTNTAVNSTNIRSFNSCTKINGDIIILRNSPEGPHYKIGTMDPEHLMNLTVEITG 407
Qy	415 YIYISAMPDLSLDLSEFONLOVIRGILNNGAYS-LTLOGLISWLGSLRLSLGLAL 473
Db	408 YLVIMWPMENMSTSLSPQNLLEIRKGTFRSGSFVYVGVRLQWGLSLKLEVSAGNYI 467
Qy	474 IHHNTHLCVHTVPMDOLEFRNPQALHTANRPEDECEGELACHOLCARGCMGPGPTQ 533
Db	468 LKNTIQLRVANTINNRRLFRSEDSQIEVART-----ENOTCNNECEDGCM-PGPTM 519
Qy	534 CYNCGQFLRGQECVEBCRVLCGLPREYVNAHCLPCHPECPQPNQSVTCFGEADQVAC 593
Db	520 CVSCCHVDRGRCVASCNLLQCEPREAOVDGRVCVCHQCLVQTDLSLTCYGPANCSKS 579
Qy	594 AHYKDPFCVACPSGVKNDLSYMPKPFDEEGACQRPINCHTSVULDDKCPAEOR 653
Db	560 AHFOGQPCITPCPCPHGILGDDPTL-IMKYADKMGQOQPHONCTQCCSGPGLSGRGD-1 637
Qy	654 ASPLTSISAVVGIILVVVLGVFGLIKRROOKIRKTYMRRLQETELVEPLTSGAMP 713
Db	638 VSHSLAVGLVSGLLITVALLIVLLRRRIK-RKRTIRCLLOEKELVEPLTSGCAP 696
Qy	714 NOQMRLIKETELRKVKVLGSGAFGTVYKGIWIPQGENYKIPVAIKVLENTSPKANKI 773
Db	697 NOAFRLIKETEFKKDRLVLSGAFGTVYKGLWNPQGENIRIPVAIKVLENTSPKANKI 756
Qy	774 LDEAVVMAGVSPVYSLIGICTSTVQVLTQMLPYGCLLDHVRERKRLGSDLLNMCW 833
Db	757 LDEAVVMASVDHPHCRLLIGICTLSAVQVLTQMLPYGCLLDVROHQRICQOMLIMCV 816
Qy	834 QIAKMSYLEDVRLVHRDLAARNVLSKPNHYKITDFGLARLLDIDETVYHADGKVPK 893
Db	817 QIAGKMYLLEERHLVHRDLAARNVLLKPNHYKITDFGLSKLLTDEKEXQADGKVPK 876
Qy	894 WMALESILRRRFTHOSDVSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRLLPQPIK 953
Db	877 WMALESIIQMTYTHOSDVSYGVTVWELMTFGSKPYDGI PAKEIASVLENGERLQPPIC 936
Qy	954 TIDVYMIWKKCMIDSECRPRELVESEPMARPOQFVYIQNDLGPASLDSTFPRS 1013
Db	937 TIEVYMIILKCMIDPSRPRELVESEPMARPSYLVYIQ--NLPSLSDRLPSR 993
Qy	1014 LLEDDMDGLVDAEELVPOQGFPCPDAPAGAGVHHRHSSSTRSGGDLTLGLPSE 1073
Db	994 LLSDD--DVVDADDEYLLPYKRI-----NRQSS----- 1019

Qy	1074 EEARSPPLAPSEAGSDVFDGLGMAKGLQSLPTHDPSPLOYSEDPV-PLPSETDG 1132
Db	1020 -----EPCIPPTGH-----PRENSITLRNLSDFQMLEKOLD 1054
Qy	1133 YVAPLTCSQPEVYNPDVRRPQ-----PSPRE-----GLP-AARPGATLERAKTL 1179
Db	1055 H-----EYVNPQGETSSRLSDIYPNYEDLTDGWPVLSQEAETNFSRPEYL 1104
Qy	1180 SPQKGVVADVAFAGVAVNPEVLTPOGGAAPRPHPPAPSPFQNLVYWDQPPRENGAP 1239
Db	1105 NTNQNSL---PLVSSSMDDPDY---QAG-----YQAF-----LPQYAL 1139
Qy	1240 PSTFKGTPTAENPEYLQ 1256
Db	1140 TQNGMFLPAENLEYLG 1156

RESULT 9  
A36223  
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 04-Oct-1991 #sequence\_revision 13-Jan-1993 #ext\_change 17-Nov-2000  
C:Accession: A36223; 159164  
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.  
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989  
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal  
A:Reference number: A36223; PMID:90083234; PMID:2687875  
A:Accession: A36223  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1342 <KRA>  
A:Cross-references: GB:M29366  
R:Plowman, G.D.; Whittney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, J.  
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990  
A:Title: Molecular Cloning and expression of another epidermal growth factor receptor  
A:Reference number: 159164; PMID:90311312; PMID:2164210  
A:Accession: 159164  
A:Status: preliminary  
A:Molecule type: translated from GB/EMBL/DBJ  
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>  
A:Cross-references: GB:M34309; NID:9183990; PIDN:AAA35979.1; PID:q306841  
C:Genetics:  
A:Gene: GDB:ERBB3; HER3  
A:Cross-references: GDB:119880; OMIM:190151  
A:Map position: 12q13-12q13  
A:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
C:Keywords: ATP; phosphotransferase  
F:707-972/Domain: protein kinase homology <KIN>  
F:715-723/Region: protein kinase ATP-binding motif

Query Match	
Beet Local Similarity	35.3%; Score 2418.5; DB 2; Length 1342;
Matches	533; Conservative 192; Mismatches 454; Indels 141; Gaps 34;
Qy	10 GLLALLPPGAA--STOVCTGDMKRLRPASPEHLDMLRHLVYCCGVYQGNLELYLPT 67
Db	11 GLLFSLANGSEVNGQAVCPGLNGLSVTGAENQYOTLYKYERCEVVMGMLLEYLTGH 70
Qy	68 NASLSFLDIOIEVQGVYLIANQVROVLPQRLRIYRGQLPEDNYALAVLDNGDPLANN 127
Db	71 NADLSFLQWIREVTSYVIVANNEFTLPLPNLRVVRGVQVYDGKFAIV-----MLANN 124
Qy	128 NPTVSFMLRVPRVNSH-LEQLR--SLTEILKGVLIQBNPOLCYODTIIMKDIFHKNNQ 184
Db	125 -----TNSSHALROLRLTQLEILISGVYIRKNDLCHMDTIDMDIVDRD- 171
Qy	185 LALTLIDTNRSRACHPCSPMCKSGSRCKGSESDCQSLTRTVACGC-ARCKGPLPTDCH 243
Db	172 --AEIVVDNDRSCPCHEVCKG-RCKPGESEDCQTLTKITCAPOCNGCHCEPNQOCH 228
Qy	244 EQCAAGCTGPKHSDCLALPHNSGIELLCPLAVTYNTDTFESMPDEGRYTFGASCVT 303



```
Db 229 DECAGGCGSRQDITDCFACRHNFDNSGACVRCFQPLVYNNKLTQLEPNPHTKYOYGGCVVA 288
QY 304 ACPYNYLSTDVSGCTLYVCPLNHNOEYTAEBGTORCEKSCSPCARVCYGLMELHREBRAVT 363
Db 289 SCPNHFV-VDTQSCVRACPPDKMEVD-KGGLKMKCECGGLCPKACGCTSG--SRFQYTD 344
QY 344 SANIOEAGCKKIFGSLAFLPESFDPPASNTAPLOPELOVEFTEETITGYLYISAMPD 423
Db 345 SSNIGFVNTCTIKLGNLDPLITGLNCDPMHKIPALDPEKLVNFRVTRLETGYLNTQSWRP 404
QY 424 SLPLDSVFONLOVIRGRILHNGAYS-LTQGLGISWGLRSJRELGSGLAIHHNTHLCE 482
Db 405 HHMNFVSFNSLTITIGRSLYNKGFSLLIMKNLVNTSLGFRSLKELISAGRIYISANQLCY 464
QY 483 VITVPMDOLEFRNPHOALHTA-NRPEDECVBEGGLACHOLCARGHCMGPPTOCVNSOCL 541
Db 465 HHSIMMTKVLKRGTEERLDIKHNRRPRDVAAGKVDPLCSSGCGGPGGGQCLSCRNYS 524
QY 542 RGOECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADOCVACAHYDPPF 601
Db 525 RGVGVCTHGNFLNGEPRFAHEACFSCHPRECQPMGATATCNGSGSDTCAQCAHPFDGPH 584
QY 602 CVARCPGCVKPLSTYPIKFPDEBACQCPPICTHSCVVDLDDKCCPAEORA---SPL 657
Db 585 CVSSCPHVLG--AKGPYIKYDPVONECRPCHENCTQCGKGPEDQLQOTLVLIQKTHL 642
QY 658 TSIVSAVVGILVVVGVVFGILIKRQOKIR-KYTMRLLOTELVEPLTPESGAMPNOA 716
Db 643 TVALVTIAG--LVIFPMGLGTFLYVRGRIRIONKRMRYLEGGSEIEPLDPS-EKANKV 699
QY 717 QMRILKETELRKVKVLSGAFGTVYKGIWIPDENVKIPVAIKVLAENTSPRANKELDE 776
Db 700 LARIFETELRLKVLGSGVFGTVHKGVMIPEGESIKIPVICIVIEDKSGRQSFQAVTDH 759
QY 777 AVYMAVGSPYVSRLLIGLICLTSTVQLVTQMLPBGCLLDHYREKRGSLGSDLLMCMQIA 836
Db 760 MLAIQSLDHAHIVRLGLCPGSSLDQVLYLPLGSLDDHYRQHRGALGQLLLMVOVQIA 819
QY 837 KCMASYLEDVRLVHRDLAARNVLVKSBNHYKITDFGLARLLIDETEVYHADGGCVPIKMA 896
Db 820 KCMYILEHGMVHRNLAAARNVLKSPQOVADFGADLLPPDKOLLSEAKTPIKMA 879
QY 897 LESILRRRTFTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 956
Db 880 LESIHGKYTHOSDVMSYGVYVWELMTFGAEPYAGLRLABVPLLEKGERLAPQICTID 939
QY 957 VYMIWVKCMIMIDSECCPRPRELVSESRAPDQRQVVIQNEBLGA---SPLDSTFYRS 1013
Db 940 VYMIWVKCMIMIDENIRTEKELANETRNAADPRYLVIKRES-GEGLAPGPEPHGLTNK 998
QY 1014 LLEDDMDGLVDAEYLVFQCGFFCDPAFGAGMWHHRHRSSTSGGDLITLGLP-S 1072
Db 999 KLEEVLEBELDLDLLEAED-----NIATTTLGSLSLPCTIAN 1039
QY 1073 EEEAPRSLAPSEGASDVFDGDLGMAKGLQSLPLETH-PSPLQRYSEDPTVPLR--- 1127
Db 1040 RPRGQSLLSPSSGY-MPMNGNLGSCOBASAVGSSERCPRVSLH-----PMRGL 1092
QY 1128 --SEIDGYVA-----PLTCSPPQE---YVNPQVTRPQPPSPREP--- 1162
Db 1093 ASESSEGHVTGSEAELOEVKSMCRSRSRSPRPDSAYHSQRHSLTLPVPLSPGLE 1152
QY 1163 -----LPARPAATLERAKTSLP-GKNGV-----KQVFAFGAVENPEYLTPOG 1207
Db 1153 EEDVNGVYVMDTHLTKTPSSREGTLLSVGLSVLTGEEBED-----EEVEYNNRR 1204
QY 1208 GAAPORPPAPSPAPDNLTYMD-----ODPBRGAPSTFYKTPAENPEYL 1255
Db 1205 RHSP-PHPRPSSLLELGYEYMDVGSLSLGSSTQSCPLHPVPIPMTAGTTPDEDEYEM 1263
```

RESULT 10  
JC4387

```
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: J04387
Gene: Heliyer, N.J.; Kim, H.H.; Greaves, C.H.; Starke, S.L.; Koland, J.G.
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protei
A:Reference number: J04387; MUID:96096555; PMID:8522190
A:Accession: J04387
A:Molecule type: mRNA
A:Residues: 1-1339 <HELV>
A:Cross-references: GB:U29339; NID:G915389; PID:G915390
A:Experimental source: liver
A>Note: The authors translated the codon AAT for residue 369 as Thr and GAT for resid
C:Comment: This protein is a functional heregulin receptor that transduces signals to
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-1339/Product: epidermal growth factor homolog #status predicted <Mat>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <Kin>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 34.1%; Score 2334.5; DB 2; Length 1339;
Best Local Similarity 40.5%; Pred. No. 1,4e-88;
Matches 523; Conservative 173; Mismatches 429; Indels 167; Gaps 36;

QY 3 LAALCRMGILLALLPFGAA---STOYCTGTDMLRLPASPEHLLDMLRLHLYOGCOVVOGN 59
Db 7 LQVLC-----FLSLARGSMGNSQAVCPGTLINGLSTGADNOYOTLLYLYKECEVVMGN 62
QY 60 LELVLYPTNASLSPFDIDIEVOGYVLIANOVQVPELQRLIRYRGTLFEDNYALAVLDN 119
Db 63 LEIVLTGNHADLSFLQWIREYAVLVANNESVLPPLMLRVVRGTVYDGKFAIFV--- 119
QY 120 GRLPLNNFNTYSPMLRVKVSASH-LEQUR--SLTEILLKGVLIQNRNOLCYQDPTLMK 176
Db 120 ---MLNLYN-----TNSSHALRLKFTQLTEILLSGVYIEKNKCLCMDDITDMR 164
QY 177 DIFHKNNQALATLIDTNRGRACHPCSPMGKSGRSGESSEDCSLTRTYCAGG--ARKG 235
Db 165 DIVRYR---GAEIVANNANPCPCHVEVKG-RMGPRGDDCOILKTTICAPCNGRCFG 220
QY 236 PLPTDCHEQCAAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRY 295
Db 221 PNPNCCHDECAAGCGGPDITDCFACRRFNDSGACVPRCPREPLVYNNKLTQLEPNPHTKY 280
QY 296 TFGASCVTAQCPNYLSTDVSGCTLYVCPLNHNOEYTAEDGTORCEKSKPCARVCYGLMEH 355
Db 281 QYGVGVVASCPIHFV-VDTQFCVRACPPDKMEVD-KGGLKMKCECGGLCPKACGCTSG- 337
QY 356 LREVRVNTSANIOEAGCKKIFGSLAFLPESFDPPASNTAPLOPELOVEFTEETITGY 415
Db 338 -SRVQVDSNSNIDGFNCTIKLGNLDPLITGLAVNDPMHKIPALDPEKLVNFRVTRLETGY 396
QY 416 LYISAMPDSLPLSVFQNLQVIRGRILHNGAYS-LTQGLGISWGLRSJRELGSGLAI 474
Db 397 LNIQSWPPIHMHFVSFNSLTITIGRSLYNRGFSLLIMKNLVNTSLGFRSLKELISAGRIY 456
QY 475 HNHTHLCFNYHTVPMOOLFNNPHOALHTA-NRPEDECVBEGGLACHOLCARGHCMGPPTO 533
Db 457 SANQDLCTHNSLNMWRLRLGPEERLDIKYDRDLGCLABGKVCDDLCSGGCGWGAPOQ 516
QY 534 CVNSQGLFAGQECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADOCVAC 593
Db 517 CLSCNYSRSEGVCTVHCNFIQGEPRFVHEADCFCHPRCLPMEGISTNGSSDCAAC 576
QY 594 AHYKQPPFCVAPCPGKPLSTYPIKFPDEBACQCPPICTHSC--VLDLDDKCCPAE 651
```



```

Db      577 AHFRGPRCVAWSCPPIIG--AKGPIYKPPADQNECRPHENCTQCCNGPELQDCLGQAE 634
Qy      652 QASPLTSIVSAVVGILLVVLGVFGILLIKRQOKIR-KYTRRLLOETELVEPTPSG 710
      635 VLMSKPHLVIAVTG--LAVILMIIGSGFLYWRGRIRQKRAMRRLLENGESIEPLDPS- 691
Qy      711 AMPNOAKRILKETELRKVKYLGSAFGTVYKGIWIPGSENVKIPAIIVLEBNTSPKAN 770
      692 EKANKVLARIFKETLRKLVKLGSGVFGTVHKGIMIPGESIKIPVICIVIDKSGROSF 751
Qy      771 KEILDEAVMAGVSGPYVSRLLGICLTSTVQLVTQMLPBGCLLDHYRENRGLSGSDLLN 830
      752 QAVTDHMLAVGSLDPAHIVRLGLCGSSLDQVLTQVPLGSLUDHVKQRETLGPPOLLN 811
Qy      831 WCMQIAKGMSTYLEDVRLVHRDLAARNVLKVSBNHYKITDFGLARLLDIDETEYHADGKY 890
      812 MGVOIAKGMVYLEHSMVHRDLARNVLMKSPSQVADPFGVADLLPPDQKLHSEAKT 871
Db      891 PIKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLP 950
      872 PIKMALESIHFGKYTHOSDWSYGVTVWELMTFGAEPYAGLRLEIPDLEKGERLAP 931
Qy      951 PICTIDVYMIWKCMMIDSECRPFRELVSFSSRMARDPQRFVVIQNEDLGASPLDSTF 1010
      932 QICTIDVYMIWKCMMIDENIRPTRELANETRKARDPPRYLVIKRAS-GEQTP--PAA 988
Db      1011 YRSLEDDMDGLVDAEYLVPOQGFCDPPAPGAGMVHHRSSSTRSGGGLTLGLE 1070
      989 EPSVLTKEL-----QEALPEL-----DLDLDE 1014
Qy      1071 PSEF-----EAPNSPLAPSE-----AGSDVFDGDLGMAAK 1102
      1015 AEEELATSLGSLSLPTGTLTRPGSOSLSPSSGYMPMNSLGEACLDAAVLGREQ 1074
Db      1103 GLQSLPTHPDPSLQRYSEDPVPLPSETDGYV---APL-----TC-----SPOPE-- 1144
      1075 FSRPISLH-PIPRGR-----PASESSGHVYTGSAELQEVSVCRSSRSRSPPRCD 1126
Qy      1145 --YVNPDPVRPOPSPSREGP-----LPAARPAAGATLERAKTSLP-SKNGV--- 1187
      1127 SAYHSQSRSLTLTPVPLSPGLEEDNGVMPDTHLRGASSRREGLTSSVGLSYLGR 1186
Db      1188 ---KDVAFAGAVENPEYLTPOGGAAPQHP 1216
Qy      1187 EEDED-----EEVEYMNKRKRGSP-PRPP 1209

```

RESULT 11

TVFVYV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C:Species: avian leukosis virus, ALV

C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999

C:Accession: B00643; A00643

R:Nilisen, T.W.; Matoney, P.A.; Goodwin, R.G.; Rottman, F.M.; Citterenden, L.B.; Raines, M. Cell 41, 719-726, 1985

A:Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and protein-tyrosine kinase in ALV-induced erythroblastosis: novel RNA processing and protein-tyrosine kinase

A:Reference number: A00643; MUID:85228222; PMID:2988784

A:Accession: B00643

A:Molecule type: mRNA

A:Residues: 1-698 <NT>

A:Cross-references: GB:M10066; GB:M13801; NID:9211749; PIDN:AAA48763.1; PID:9211750

C:Note: in Genbank entry CHKRBFR, release 109.0, the source is designated as Gallus gallus

C:Comment: This protein is synthesized as a gag-env-erbB protein.

C:Genetics:

A:Gene: gag-env-erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P

F:1-6/Product: gag protein (fragment) #status predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>

F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ENK>

F:194-459/Domain: protein kinase homology <KIN>

F:202-210/Region: protein kinase ATP-binding motif

F:229/Active site: Lys #status predicted

Query Match 25.8%; Score 1766.5; DB 1; Length 698;

Best Local Similarity 52.2%; Pred. No. 1,4e-65;

Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

```

Qy      584 GPEADQCVACAHYKDPFCVAPCPGCVKPDLSYMPIWKPPEDEGACOPCEINCHSCVDL 643
      60 GP--DHCMCAFIIDGPHCVKAPGAVLGENDTL--VMKADANAACVQLCHPNTCRGCKP 116
Db      644 DDKCPAEORASPLTSIVSAV--GILLVVLGVFGILLIKRQOKIRKTYTRRLLOETEL 702
      117 GLEGGP--NGSKTBSIAGVVGILLVVGIGLYLRR--HYRKTRRLLOEREL 172
Qy      703 VEPLTPSGAMPQAKRILKETELRKVKYLGSAFGTVYKGIWIPGSENVKIPAIIVLEBNT 762
      173 VEPLTPSGAMPQAKRILKETELRKVKYLGSAFGTVYKGIWIPGSENVKIPAIIVLEBNT 762
Db      763 ENTPSPKANKIIDEAVMAGVSGPYVSRLLGICLTSTVQLVTQMLPBGCLLDHYRENRGR 822
      233 EATSPKANKIIDEAVMAGVSGPYVSRLLGICLTSTVQLVTQMLPBGCLLDHYRENRGR 292
Qy      823 LGSODLWCMQIAKGMSTYLEDVRLVHRDLAARNVLKVSBNHYKITDFGLARLLDIDETE 882
      293 IGSOYLWCMQIAKGMSTYLEDVRLVHRDLAARNVLKVSBNHYKITDFGLARLLDIDETE 352
Db      883 YHADGKVPDKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLE 942
      353 YHADGKVPDKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLE 412
Qy      943 KGERLPQPICTIDVYMIWKCMMIDSECRPFRELVSFSSRMARDPQRFVVIQNEDLGAS 1001
      413 KGERLPQPICTIDVYMIWKCMMIDADSRPFRELIAFSKARDPPTYLVIQGERMH 472
Qy      1002 PASPLDSTYRSLDEDDMDGLVDAEYLVPOQGFCDPPAPGAGMVHHRSSSTRSG 1061
      473 LPSPTDSKRYRILMEEDMEDIVDAEYLVPOQGF-----NSPST--- 513
Db      1062 GGDLTGLLEPSEEAAPRSL-----APSEAGSDVDVDGLGMAKGLSLPTHPDPSPLQ 1116
      514 -----SRITLLSSLSATSNSTNCID-----RNGGHPREDSFVQ 550
Qy      1117 RYSEDPVPLPSET--DGVAALTCSPQPEYVNPDPVRPOPSPREGPLPAARPAATLE 1174
      551 RYSDPTGFLRESIDGFL-----PAPEYVNO--LMPKKRS----- 585
Db      1175 RAKTSLSPKNGVYKVF-----AFGAVENPEYLTPOGGAAPQHP 1219
      586 ---TAMVONQIYNNISLTAISKLPWDSRYQNSHTAVDNPETL-----NTNOSPULA 633
Qy      1220 SPAPNLYWQD-----DPPE-----RGAPSTKGTPTAENPEYLGIDVP 1260
      634 KTVFESSPYWIOSGNHQLNDPNPYQDPLPNETYKNGLLKVPAAENPEYLRVAAP 689

```

RESULT 12

TVYVUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)

C:Species: avian erythroblastosis virus

C:Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999

C:Accession: A00644; A38022

R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K. Cell 35, 71-78, 1983

A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family

A:Reference number: A00644; MUID:84026539; PMID:6133229

A:Accession: A00644

A:Molecule type: DNA

A:Residues: 1-604 <YAM>

A:Cross-references: GB:K01216; NID:9209676; PIDN:AAA42400.1; PID:9209678

R:Debnire, B.; Henry, C.; Benaissa, M.; Bisette, G.; Claverie, J.M.; Saulie, S.; Martin

Science 224, 1456-1459, 1984

A:Title: Sequencing the erbB gene of avian erythroblastosis virus reveals a new type of

A:Reference number: A38022; MUID:84223957; PMID:6328658

A:Accession: A38022  
 A:Molecule type: DNA  
 A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>  
 A:Cross-references: GB:K02006  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P  
 F:130-395/Domain: protein kinase homology <KIN>  
 F:138-146/Region: protein kinase ATP-binding motif  
 F:165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;

Best Local Similarity 52.2%; Pred. No. 4, 8e-63;  
 Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```

Qy 593 CAHYKDPFCVAPGSGVPRDLSPYMPKPEDEGACQPCPINCYSQVLDKDCPAAEQ 652
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VRKYADANAVCOLCHPNCCTGCKGPGLEGP--- 58
Qy 653 BASPLTSISAVY-GILAVVVLGVVGIILKRRQKIKRYTMRLLQETELVEPLTPSGA 711
Db 59 NSGKTPSIAGVVGGLCLVVGVLGILVLRK-HIVRKKTLLRLQERLVEPLTPSGE 117
Qy 712 MPNOAQMRILKETELKRVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVRENTSPKANK 771
Db 118 APRQALRLIKETEFKVKVLSGAGFTIYKGIWIPDGENVKIPVAIKVRENTSPKANK 177
Qy 772 EILDEAYVAVGSPYVSRLIGLITSTVOLVTLQMPYGLLDHVRNRRGLCSODLLNM 831
Db 178 EILDEAYVAVSVNPHVCRLIGLITSTVOLVTLQMPYGLLDHVRNRRGLCSODLLNM 237
Qy 832 CMQIAGMSYLEVRLVHRLDARNVLYKSPNHYKTTDECLADLLDDEHYADGKVP 891
Db 238 CVOIAGMNYLERRLVHRLDARNVLYKTPQVHKITDFGLAKLIGADEHYAEGKVP 297
Qy 892 IKMALESILRRFTQSDVMSYGVTVMEIMTFGAKPYDGIIPAREIPDLLEKGERLPQRP 951
Db 298 IKMALESILRRFTQSDVMSYGVTVMEIMTFGSKYDGIIPAREIPDLLEKGERLPQRP 357
Qy 952 ICTIDYVMIMVKCMIDSECRPRRELVSFSPMARDPQRFVVIQ-NEDLGASPLDSTF 1010
Db 358 ICTIDYVMIMVKCMIDSECRPRRELVSFSPMARDPQRFVVIQ-NEDLGASPLDSTF 417
Qy 1011 YRSLDEDDMGDLVDAEYVLVPOGFCPPDAPAGAGMWHRRSSSTRSGGDLTLGLE 1070
Db 418 YRSLDEDDMGDLVDAEYVLVPOGFCPPDAPAGAGMWHRRSSSTRSGGDLTLGLE 449
Qy 1071 PSEBEAPRSPPL-----APSEAGSDVDFDGLGMAKGLQSLPTHDPSPLOQRYSEDPV 1125
Db 450 -----SRPLLSLSLSTSNNSATNCID-----RNGGHPVREDSFVQRYSSDPTGN 495
Qy 1126 LPSET--DGVAVALTSGPOEYVNOQRPVQRPSPRGRPLPAARFAGATLERAKTSLPGK 1183
Db 496 PLEESIDGFL-----PAPEYVNO--LMPKRPSTAN----- 524
Qy 1184 NGVVKOVFAF-----GAVENPEYLTTPGGAAPQHPAPAFSPAFD 1224
Db 525 --VQNOIYFISLTALSKLPMDSRYQNSHSTAVDNPEYL-----NTQSPFLAKTVFE 574
Qy 1225 NLVYWDQDPPERGAPSPSTFGKTPAENPEY 1254
Db 575 SSPYWIQSGNHQ-----INLDNPVY 594

```

# RESULT 13

GOPE  
 epidermal growth factor receptor - fruit fly (Drosophila melanogaster)  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: Drosophila melanogaster  
 C>Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 11-Jun-1999  
 C:Accession: A00640; A38021  
 R:Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.

Cell 40, 599-607, 1985  
 A:Title: The Drosophila EGF receptor gene homology: conservation of both hormone bindi  
 A:Reference number: A00640; MID:85124611; PMID:2382499  
 A:Molecule type: DNA  
 A:Accession: A00640  
 A:Residues: 1-1330 <Liv>  
 A:Cross-references: EMBL:K03054  
 R:Madsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.  
 Nature 314, 178-180, 1985  
 A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor  
 A:Reference number: A38021; MID:85137938; PMID:2383232  
 A:Accession: A38021  
 A:Molecule type: DNA  
 A:Residues: 'A', 832-866, 'V', 868-943, 'QTPSLYK' <MAD>  
 A:Cross-references: EMBL:X02293; MID:g7922; PIN:CAA26157.1; PID:g929565  
 C:Comment: This sequence is tentative because the introns have not been identified.  
 C:Genetics:  
 A:Gene: FlyBase:Bgfr  
 A:Cross-references: FlyBase:Fgdn0003731  
 A:Map position: 2 57F  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos  
 F:1-732/Domain: extracellular #status predicted <EXT>  
 F:733-764/Domain: transmembrane #status predicted <TM>  
 F:765-1330/Domain: intracellular #status predicted <INT>  
 F:808-1072/Domain: protein kinase homology <KIN>  
 F:816-824/Region: protein kinase ATP-binding motif  
 F:122-300,324,363,518,688,695,700/Binding site: carbohydrate (Aen) (covalent) #status  
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predi  
 F:843/Active site: Lys #status predicted  
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pred

Query Match 24.2%; Score 1655.5; DB 1; Length 1330;

Best Local Similarity 29.9%; Pred. No. 9, 2e-61;  
 Matches 414; Conservative 180; Mismatches 414; Indels 377; Gaps 40;

```

Qy 80 VQGYVLIANQVQVPLQRIYRGTOF-----EDVYALAVDNGDPLNFFNFTGFW 134
Db 38 ITMYIYIGDLICTLSYRIQIRGLFSLSEBEKAL-----FVYSKMYT 86
Qy 135 LRVKVASASHLEQLRSITELIKGCVLIQRPOLCYODTILMKDI FHKNOALATLIDTNR 194
Db 87 LEIP-----DLRVLNQGQVGFHNHYNLCHMRTIQMSEIVSNGTDAVYVDTAP 135
Qy 195 SRACHCSPKCKSRCKMGESSEDCQSLTRVACAGCA--RCKGLPDPCCHEGCAACTG 252
Db 136 ERECPKCHESCTHG-CMGEGPKNCQKFSKLTCSPPQCGAGRCYGRPRECHLFCAGGCTG 194
Qy 253 PKASDCLACLHFNHSGICELHCPALVYNTDFESMNPREGRYTFGASCATACPYNYLST 312
Db 195 PLOKDCIACGNFDEAVSKCECPMKRYNTTTLVLEINPBGKAYGATCYKECP-GHLLR 253
Qy 313 DVGSCVLCPHLNHOEYTAEDGTORCEKSPKACVCGYGLMEHLREYRAVTSANIOEPAG 372
Db 254 DNGACVRSQPDQKMDKGE-----CVPCNGPCPKTCGTVLH-----AGNIDSFN 300
Qy 373 CKKIFGSLAPSPFG--DPAANTA-----PLOPQLOVFELEITGYLYISAMPDS 424
Db 301 CTVIDNIRILDTDFSGFQDVVANYTWGPRYIPLDPERRNVFSTVKEITTYLLENGTHPQ 360
Qy 425 LPLSVFQNLQVIRGLIHNGAY-SLTQIGISWILRLREISGLALIHNTHLCFV 483
Db 361 FRNLSTFRNLLETIHGQLMSPALAIIVSSLSLEMRDLKQISSGVVYIQHNRDLCTV 420
Qy 484 HTVPMDQLFRNPQALLHTANRDEDC----- 510
Db 421 SNIRMPAIQKEPQKQVNVNENLRADLCGKFLITLISVQHNIIIMHIFAIKREKNHLLGSV 480
Qy 511 ----- 510
Db 481 QRGRLLSWHSVPLVLOELQFQWHLHRLMLYIQVINSITQDSNEHQLTDACYSPSVPT 540
Qy 511 -----VG 512

```

Db 541 SUTIRARAYAIOSAGLAMELEBOTIARSASMRHSKITPAEGROVPRWVELGVCASARAGIA 600  
Qy 513 EGLA-----CHOLCARGHCWGPFTQCVCNSQFLRGQCEVEECVLOGLPREVY---N 562  
Db 601 EPLABRAVCRKCHPLCELCTNYGHNQVSKCHYRRRQCEEC-----PADHTDDE 654  
Qy 563 ARHCLPCHPECOPOGNSVTCFPGPADQVACAHYK-----DPPF-----CVARCPGG 609  
Db 655 .ORECQORHPEC---NG---CTGPGADDCSKSCNFKLPDANETGPPYNSITMFNCTSKCPLE 708  
Qy 610 VK-PRLSYWPIWKFPDEGACOPCPINCHSCVDLDDKDCPABGRASPLTSIVASAVGL 668  
Db 709 MNAVYQTAICPY-----CAASPPRSSKITANLD-----VNMFTITYGAV 749  
Qy 669 LVVVLGVVFGI-LIKRQOKIRRYT--MRRLQETELVEPLTPSGAMPNOQMRILKETE 725  
Db 750 LVPTICILCVTYTICRQKAKKETVKMTMALSGRDSPLRPSNIGANLCTRIYKDAE 809  
Qy 726 LKRVKVLGSGAGFYTKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVMAGVS 785  
Db 810 LRKGVLLGWFAGFGRVYKGVWPEGENVKIPVAIKELKSTGAESEEFREAYIMASEEH 869  
Qy 786 PVSRLIGICTSTVQLVTQLMPYGLDHLVRENHGRGLSODLLNMCQIAGKSYLDEV 845  
Db 870 VNLKLLAVCWSQMLLTQLMPLGLDLYVRNNRDKISKALLNMTQIAGMSYLEEK 929  
Qy 846 RLVRHDLAARNLVK---SPNHVKITDFGLARLLDIDETEHADGKVPDKMALESILR 902  
Db 930 RLVRHDLAARNLVLLAGEDI---DFGLAKLLSDSNEMKYKAAGKMPKIMLALECIKN 985  
Qy 903 RRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGBRLPQPICTIDVYMINV 962  
Db 986 RVFTSKSDVMAGVITWELTFGQRPHEHNPAKDIPDLIEVGIKLEQPEICSLDIYCTL 1045  
Qy 963 KCMWIDSECRPFRELVSFESRMARDPQRFVVIQNEDELG--PASPLDSFFYSLEDD-- 1018  
Db 1046 SCHMIDLAAMRPTFKQTLTFAFAPARDPGRYLAIIIGSKTRLPK-----YTSQDKDLI 1098  
Qy 1019 -DMGDLVDAEYLVPQGFPCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEEAR 1077  
Db 1099 RKLAPTTDGSSEAIAPRDYLGKALGPS-----HRTDCT-----DENP 1137  
Qy 1078 -----RSPLASSEAGSGSVFDG---DLGMAKAKGIQSLPTHDPSPFLQYSDDPYVLP 1128  
Db 1138 KLNRYCKDPSNKNSSSTGDERDSSAREVGVNLR-----LDLPV 1176  
Qy 1129 ETDGVAVPLTCSPOPEYVNPQDVRPQPSPREGPIPAARPAAGATLERAKTISPKGNKVYK 1188  
Db 1177 DEDDYLP-TCOPGPRNNNNNN-----NPNQNNMAAGVAAAGYM----- 1214  
Qy 1189 DVFAFGAVENDEYL-----TPOGGAAPQPH-----PPAFSP-APDN 1225  
Db 1215 DLIGVPSVDNPEYLLNQTGLGVGESPIPTQITGIPVMGPGTMEVKVYMPQSEPRSSDH 1274  
Qy 1226 LYIWD 1230  
Db 1275 EYIWD 1279

## RESULT 14

S35745  
C:Species: avian erythroblastosis virus  
C:Date: 03-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 26-Feb-1997  
C:Accession: S35745  
R:Vennstrom, B.  
Submitted to the EMBL Data Library, March 1993  
A:Reference number: S35743  
A:Accession: S35745  
A:Molecule type: DNA  
A:Residues: 1-544 <VEN>  
A:Cross-references: EMBL:X12707

C:Genetics:  
A:Gene: erbB  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific  
F:135-400/Domain: protein kinase homology <Kin>  
F:143-151/Region: protein kinase ATP-binding motif  
F:170/Active site: Lys #catalue Predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;  
Best local similarity 54.9%; Pred. No. 8, 6e-61;  
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 584 GPEADQVACAHYKDPFVVARCPGCVKPDLSYMPIWKPDEEGACOPCPINCHSCVDL 643  
Db 1 GP--DHCMCAHFIDDPHCVCACAPAGVLGENDTL-VWKADANAVQCLCHPCTRCCKGP 57  
Qy 644 DDKCGPABRASPPLTSIVSAVY-GILVYLVGVFPIILIKRQOKIRKTMRLLODEL 702  
Db 58 GLEGP--NGSKTPSIAGVVGGLCLVAVGIGLGLYLRR-HIVRKTLRLLOREL 113  
Qy 703 VEPLTPSGAMPNOQMRILKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVL 762  
Db 114 VEPLTPSGAMPNOAHRLIKETEFKVKVLSGAGFTYKGLMIPGEKVTIFVAIKEL 173  
Qy 763 ENTSPKANKELIDEAYVMAGVSPYVSRLLGICTSTVQLVTQLMPYGLDHLVRENHGR 822  
Db 174 EATSPKANKELIDEAYVMAGVSNPNHVCRLIGICTSTVQLVTQLMPYGLDYLIRHKDN 233  
Qy 822 LGSODLLNMCQIAGKSYLDEVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETE 862  
Db 234 IGSOYLLNMCVOIAGKMYLERHVRDLAARNLVKTPQHKITDFGLAKOLGADKE 293  
Qy 883 YHADGKVPDKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDL 942  
Db 294 YHAEKGKVPDKMALESILHRYTHQSDVMSYGVTVWELMTFGSKYDGI PASEISSVLE 353  
Qy 943 KGERLPQPICTIDVYMINVVKCMWIDSECRPFRELVSFESRMARDPQRFVVIQ-NEDELG 1001  
Db 354 KGERLPQPICTIDVYMINVVKCMWIDSECRPFRELVSFESRMARDPQRFVVIQ-NEDELG 1001  
Qy 1002 PASPLDSFFYSLEDDGDLVDAEYLVPQGFPCDPAPAGAGVHHRRSSSTRSG 1061  
Db 414 LPSPLDSKFRYTLMEEDMEDIVDAEYLVPQGF-----NSPST--- 454  
Qy 1062 GGDLLTGLEPSEEARSPPL-----APSEGASDVVDGLGMAKAKGLOSPLTHDPSPLQ 1116  
Db 455 -----SRPLLSLSLTSNNSATNCLDRNG-----H----- 481  
Qy 1117 RYSEDPVPLPSETDGYAPLTCSPOPEYVNPQDVRPQPSPREGPIPAARPAAGAT-LER 1175  
Db 482 -----PVREDOFL-----PAPEYVNG--LMPKKRSTAMVQNGQYNYISLTAISK 523  
Qy 1176 AATLSPKGNKVYKDVFAFGAVENDEYL 1203  
Db 524 LPIDSRYN-----SHSTAVDNPEYL 544

## RESULT 15

S00727  
C:Species: avian erythroblastosis virus  
C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 26-Feb-1997  
C:Accession: S00727  
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.  
Oncogene Res. 1, 265-278, 1987  
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mut.  
A:Reference number: S00727; MUID:88217326; PMID:2897102  
A:Accession: S00727  
A:Molecule type: DNA  
A:Residues: 1-545 <SCO>  
A:Cross-references: EMBL:X06943  
C:Genetics:  
A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; phosphotransferase  
F:135-400/Domain: protein kinase homology <KIN>  
F:143-151/Region: protein kinase ATP-binding motif

Query Match 24.0%; Score 1640; DB 2; Length 545;

Best Local Similarity 54.9%; Pred. No. 1.7e-60;  
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

```
QY 584 GREADQVACAHYKDPFCVARGSPGVKPDLSYMPIKFPDEGACQPCPTNCHSCVDL 643
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GP--DHCMKCAHFDGPHCVACAPAGVAGENDTL-WMKYADANAVCQLCHPCTGCKGP 57
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 644 DKGCPAEORASPLTSIVSAV--GILLVVVLGVVFGILIKRQOKIRKYTRRLQETEL 702
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 GLECCP---NGSKTPSIAGVYGGGLCLVVGGLGIGLYLR--HIVKRTIARLLQEREL 113
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 703 VEPLTPSGAMPNQAKMRLKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKYLR 762
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 VEPLTPSGEAPNQAHRLKETEFKVKVILGFGAFGVYKGLWIPGEKVTIPVAIKELR 173
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 763 ENTSPKANKELDEAYVMAGVSPYSRILGICLTSTVQLVTOLMPYGCULDHYRENRGR 822
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 174 EATSPKANKELDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCULDHYREHKN 233
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 823 LGSODLLNMCQIAKGNSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDIDETE 882
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 234 IGSQYLLNMCQIAKGNNTYLERLVRDLAARNVLKTPQDVKITDFGLAKOLGADKE 293
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 883 YHADGCVPTIKMALESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLE 942
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 YHAEGGVPTIKMALESILHRIYTHOSDWSYGVTVWELMTFGSKPYDGIIPASEISYLE 353
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 943 KGERLPOPICTIVVMIMVCMWIDSECRPRELVESESRMARDPQRFVLIQ--NEDLG 1001
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 354 KGERLPOPICTIVVMIMVCMWSDSDSRPKFRELIAEFSKMARDPPEYLVLIQGERMR 413
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1002 PASPLDSTFYRSLDEDDMDLVDAEELYVPOGFFCPDPAPAGAGWVHHRSSSTRSG 1061
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 414 LPSPTDSKFYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST--- 454
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1062 GGDLLTLGLPSEEEAPRSPL-----APSEGASDVFPDGLMGAKLQSLPTHDPSPLO 1116
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 455 -----SRTPLLSLSLATSNNSATNCIDRMNG-----H----- 481
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1117 RYSEDPVPLPSETDGVVAPLTCSPOPEYVNPQVVRPOPSPREGPLPAARPAQAT--LER 1175
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 482 -----PVRDEGFL-----PAREYVNO--LMPKKPSTAMVONQIYVYISLTAISK 523
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1176 AKTLSPGKNGVKKDVFAFGAVENPEYL 1203
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 524 LPMDSRYON-----SHSTAVDNPEYL 544
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 22, 2003, 09:08:27  
Job time : 32.0157 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)  
5088.033 Million cell updates/sec

Title: SEQ4-103-117-14

Perfect score: 6847  
Sequence: 1 MELALCRWGLLLALPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6708	98.0	1255	1	ERB2_HUMAN
2	5946	86.8	1257	1	ERB2_RAT
3	5921.5	86.5	1254	1	ERB2_MESAU
4	3137	45.8	1210	1	EGFR_HUMAN
5	3109	45.4	1210	1	EGFR_MOUSE
6	2986.5	43.6	1308	1	ERB4_HUMAN
7	2967	43.3	1308	1	ERB4_RAT
8	2695.5	39.4	1167	1	XMRK_XIPMA
9	2427.5	35.5	1342	1	ERB3_HUMAN
10	2357.5	34.4	1339	1	ERB3_MOUSE
11	1977	28.9	1426	1	EGFR_DROME
12	1749.5	25.6	634	1	ERBB_ALV
13	1703	24.9	604	1	ERBB_AVIER
14	1630	23.8	540	1	ERBB_AVIRU
15	1594	23.3	703	1	EGFR_CHICK
16	1296	18.9	1323	1	L723_CAEBL
17	1142.5	16.7	245	1	ERB2_MOUSE
18	727	10.6	1363	1	ILPR_BRALA
19	706	10.3	1382	1	INSR_HUMAN
20	701	10.2	1383	1	INSR_MOUSE
21	700.5	10.2	1372	1	INSR_MOUSE
22	697	10.2	1300	1	IRR_MOUSE
23	687	10.0	1297	1	IRR_HUMAN
24	683.5	10.0	1300	1	IRR_MOUSE
25	681	9.9	1607	1	MLP7_LYMS
26	671	9.8	1477	1	HTK7_HYDAT
27	640	9.3	1367	1	IGIR_HUMAN
28	630	9.2	1373	1	IGIR_MOUSE
29	626.5	9.1	1390	1	INSR_AEDAE
30	616	9.0	1390	1	INSR_HUMAN
31	610	8.9	2146	1	INSR_DROME
32	597	8.7	987	1	EPB4_MOUSE
33					

34	595.5	8.7	902	1	EPBB_XENLA	Q91736 xenopus lae
35	594	8.7	984	1	EPB1_RAT	P09759 ratius norv
36	592	8.6	985	1	EPBA_XENLA	Q91571 xenopus lae
37	591.5	8.6	984	1	EPB1_CHICK	Q07494 gallus gall
38	589.5	8.6	977	1	EPB2_MOUSE	Q03145 mus musculu
39	588	8.6	984	1	EPB1_HUMAN	P54762 homo sapien
40	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
41	584.5	8.5	976	1	EPB2_HUMAN	P29317 homo sapien
42	576	8.4	986	1	EPB3_CHICK	Q07498 gallus gall
43	575.5	8.4	998	1	EPB3_HUMAN	P54753 homo sapien
44	573.5	8.4	1053	1	FAK1_CHICK	Q00944 gallus gall
45	569	8.3	1068	1	FAK1_XENLA	Q91738 xenopus lae

## ALIGNMENTS

```

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; Pubmed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Sema K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.",
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; Pubmed=299974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Liberman T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.",
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; Pubmed=2995967;
RA Sema K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.",
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; Pubmed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.",
RL Genomics 15:426-429(1993).
-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
ALPHA AND AMPHIREGULIN.
-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
-1- SUBCELLULAR LOCATION: Type I membrane protein.

```

CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC RESIDUES (BY SIMILARITY).  
CC -!- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M11767; AAA35808.1; -.  
DR EMBL; M11761; AAA35808.1; JOINED.  
DR EMBL; M11762; AAA35808.1; JOINED.  
DR EMBL; M11763; AAA35808.1; JOINED.  
DR EMBL; M11764; AAA35808.1; JOINED.  
DR EMBL; M11765; AAA35808.1; JOINED.  
DR EMBL; M11766; AAA35808.1; JOINED.  
DR EMBL; M11730; AAA5493.1; -.  
DR EMBL; M12036; AAA5978.1; -.  
DR EMBL; X03363; CAA27060.1; -.  
DR PIR; A25491; A25491.  
DR PIR; A24571; A24571.  
DR HSSP; P11352; 1FGK.  
DR Genem; HGNC:3430; ERBB2.  
DR MIM; 164870; -.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR InterPro; IPR004019; YLP\_mocif.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recept\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR Prodom; PD000001; Euk\_kinase; 1.  
DR SMART; SM00261; Fu; 3.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KM Polymorphism.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
FT TRAASMEN 653 675 POTENTIAL.  
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 720 987 PROTEIN KINASE.  
FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
FT BINDING 753 753 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 220 227 BY SIMILARITY.  
FT DISULFID 224 235 BY SIMILARITY.  
FT DISULFID 236 244 BY SIMILARITY.  
FT DISULFID 240 252 BY SIMILARITY.  
FT DISULFID 255 264 BY SIMILARITY.  
FT DISULFID 268 295 BY SIMILARITY.  
FT DISULFID 299 311 BY SIMILARITY.  
FT DISULFID 315 331 BY SIMILARITY.  
FT DISULFID 334 338 BY SIMILARITY.  
FT DISULFID 511 520 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 531 540 BY SIMILARITY.  
FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
FT DISULFID 567 584 BY SIMILARITY.  
FT DISULFID 587 596 BY SIMILARITY.  
FT DISULFID 600 623 BY SIMILARITY.  
FT DISULFID 626 634 BY SIMILARITY.  
FT DISULFID 630 642 BY SIMILARITY.  
FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 654 654 I -> V.  
FT VARIANT 655 655 /FTID=VAR\_004077.  
FT VARIANT 655 655 I -> V.  
FT VARIANT 655 655 /FTID=VAR\_004078.  
FT CONFLICT 1170 1170 P -> A (IN REF. 2).  
SQ SEQUENCE 1255 AA; 137909 MM; 39E9DFDA04DCF962 CRC64;

Query Match 98.0%; Score 6708; DB 1; Length 1255;  
Best Local Similarity 98.1%; Pred. No. 0;  
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELAALCRWGLLALLPRGAASVOVCTGDMKRLRLPASBETHDMLRHLVYOGGVQGNL 60  
DB 1 MELAALCRWGLLALLPRGAASVOVCTGDMKRLRLPASBETHDMLRHLVYOGGVQGNL 60  
QY 61 ELTYLPTNASLFLQDIOGVGVYLAHNOVROVPOQRIRIVNGTOLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLFLQDIOGVGVYLAHNOVROVPOQRIRIVNGTOLFEDNYALAVLDNG 120  
QY 121 DPLNPNFTVSFWLKVPRVSAAS---HLEQLSLTEILKGVYLIQRNPOLCYODTILMK 176  
DB 121 DPLNPNFTVSFWLKVPRVSAAS---HLEQLSLTEILKGVYLIQRNPOLCYODTILMK 176  
QY 121 DPLNPNFTVSFWLKVPRVSAAS---HLEQLSLTEILKGVYLIQRNPOLCYODTILMK 176  
DB 121 DPLNPNFTVSFWLKVPRVSAAS---HLEQLSLTEILKGVYLIQRNPOLCYODTILMK 176  
QY 177 DIFHKNNOLALTIDNBRACHPCSPMKGSGCWESSEDDCSLRTVACGACACCKGP 236  
DB 177 DIFHKNNOLALTIDNBRACHPCSPMKGSGCWESSEDDCSLRTVACGACACCKGP 236  
QY 237 LPTDCHEQCAAGCTGPKASDCLAHFNHSGICEHCPALVYNTDTFESMNPGRYT 296  
DB 237 LPTDCHEQCAAGCTGPKASDCLAHFNHSGICEHCPALVYNTDTFESMNPGRYT 296  
QY 231 LPTDCHEQCAAGCTGPKASDCLAHFNHSGICEHCPALVYNTDTFESMNPGRYT 290  
DB 231 LPTDCHEQCAAGCTGPKASDCLAHFNHSGICEHCPALVYNTDTFESMNPGRYT 290  
QY 297 FGASCTYACPVNLTSDVSGCTLVCPHNOEVTAEADGTORCEKSKPCARVCGLGMEHL 356  
DB 297 FGASCTYACPVNLTSDVSGCTLVCPHNOEVTAEADGTORCEKSKPCARVCGLGMEHL 356  
QY 357 REVRANTSANIOEFACCKKIFGSLARLPSSFGDDPASNTAPLOPELOVFEETLEETGYL 416  
DB 357 REVRANTSANIOEFACCKKIFGSLARLPSSFGDDPASNTAPLOPELOVFEETLEETGYL 416  
QY 417 YISAMPDSLPLDSVFNOLQVIRGRILHNGAYSLLTGLGISWLGSLRELSSGLALIH 476  
DB 417 YISAMPDSLPLDSVFNOLQVIRGRILHNGAYSLLTGLGISWLGSLRELSSGLALIH 476  
QY 411 YISAMPDSLPLDSVFNOLQVIRGRILHNGAYSLLTGLGISWLGSLRELSSGLALIH 470  
DB 411 YISAMPDSLPLDSVFNOLQVIRGRILHNGAYSLLTGLGISWLGSLRELSSGLALIH 470  
QY 477 NTHLCFVHTVPMDOLEFRNPHOALLHTANRPEDECEVEGLAQHLCARGHGWGPQCVN 536  
DB 477 NTHLCFVHTVPMDOLEFRNPHOALLHTANRPEDECEVEGLAQHLCARGHGWGPQCVN 536  
QY 537 CSQPLRGQCEVECEVLOGLPREVYNARCLCHPECCQONGSVTFEGEADQVACAY 596  
DB 537 CSQPLRGQCEVECEVLOGLPREVYNARCLCHPECCQONGSVTFEGEADQVACAY 596  
QY 597 KQPPCVARCPGAVPDLSSYMPIMKFPDEGACQPCPINCTSHCVLDLKGCPAERASP 656  
DB 597 KQPPCVARCPGAVPDLSSYMPIMKFPDEGACQPCPINCTSHCVLDLKGCPAERASP 656  
QY 657 LTSISAVVGIILVVVLGVVFGILIKRROQKIRKTYMRRLQETELVEBLTPSGAMPND 716  
DB 657 LTSISAVVGIILVVVLGVVFGILIKRROQKIRKTYMRRLQETELVEBLTPSGAMPND 716  
QY 651 LTSISAVVGIILVVVLGVVFGILIKRROQKIRKTYMRRLQETELVEBLTPSGAMPND 710  
DB 651 LTSISAVVGIILVVVLGVVFGILIKRROQKIRKTYMRRLQETELVEBLTPSGAMPND 710



FT DISULFID 517 530 BY SIMILARITY.  
 FT DISULFID 533 542 BY SIMILARITY.  
 FT DISULFID 546 562 BY SIMILARITY.  
 FT DISULFID 565 578 BY SIMILARITY.  
 FT DISULFID 589 585 BY SIMILARITY.  
 FT DISULFID 589 598 BY SIMILARITY.  
 FT DISULFID 602 625 BY SIMILARITY.  
 FT DISULFID 628 636 BY SIMILARITY.  
 FT DISULFID 632 644 BY SIMILARITY.  
 FT MOD\_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1257 AA, 138831 MM, 6129264583011402 CRC64.

Query Match 86.8%; Score 5946; DB 1; Length 1257;  
 Best Local Similarity 87.1%; Pred. No. 1,8e-311;  
 Matches 1099; Conservative 50; Mismatches 107; Indels 6; Gaps 3;

QY 1 MELALCRWGLLALIPGAASTQVCTGIDMKLRLPASPETHLDMRLHYQGCQVVGNL 60  
 DB 1 MELAACRMGFLALLPFGIAGTQVCTGIDMKLRLPASPETHLDMRLHYQGCQVVGNL 60  
 QY 61 ELTYLTNASTLFDIOEYGVVLLAHNOVROVRLRIYVGTQLFEDNYALAVDNG 120  
 DB 61 ELTYLTNASTLFDIOEYGVVLLAHNOVROVRLRIYVGTQLFEDNYALAVDNG 120  
 QY 121 DPLANNFTVFSFWLVPKVASHLLEQLSLTEILKGVLLIQRPOLCYODTILMKDIFH 180  
 DB 121 DPLANNFTVFSFWLVPKVASHLLEQLSLTEILKGVLLIQRPOLCYODTILMKDIFH 180  
 QY 121 DPGDNVAASRPG--RTP--GLRELQSLTEILKGVLLIRNPOLCYODTILMKDIFH 175  
 DB 121 DPGDNVAASRPG--RTP--GLRELQSLTEILKGVLLIRNPOLCYODTILMKDIFH 175  
 QY 181 KXNQLALTLIDTRSRACHPSCMKSGRCWGESSEDCQSLRTVCAGSGARCKPLPTD 240  
 DB 176 KXNQLAPVLDITRSRACPCAPACDKMCMGESPECCQILGTIGICSGARCKGLPTD 235  
 QY 241 CCHEOCAAGCTGPKHSDCLAFHNSGICELHCPALVTNTPTFSPMPREGRTYFGAS 300  
 DB 236 CCHEOCAAGCTGPKHSDCLAFHNSGICELHCPALVTNTPTFSPMPREGRTYFGAS 295  
 QY 301 CVTACPYNYLSTDVSGCTLVCPPLHNOEVTAEQTORCEKSKPCARVCYGLGMEHLREYV 360  
 DB 296 CVTTCYNYLSTEVSGCTLVCPPNNOEVTAEQTORCEKSKPCARVCYGLGMEHLREYV 355  
 QY 361 AYTSAIOEPAGCKTIFGSLAFIPESFDGPASNTAPLOEQLQVETLEETIGYLYISA 420  
 DB 356 AITSDVVOEFDDCKTIFGSLAFIPESFDGPSSGIALPRLPEQLQVETLEETIGYLYISA 415  
 QY 421 WPDSLPDLVFOQLVIRGRILHNGAYSLTLOGLIGISWGLRSRLRELSGLALIHNTTH 480  
 DB 416 WPDSLPDLVFOQLVIRGRILHNGAYSLTLOGLIGISWGLRSRLRELSGLALIHNTTH 475  
 QY 481 CFVHTVPMQDLFRNPQALLHTANRPDE- CVEGELACHOLCARGHCKWGPPTQCVNCSQ 539  
 DB 476 CFVHTVPMQDLFRNPQALLHTANRPDE- CVSSGLVCHNGHCKWGPPTQCVNCSH 535  
 QY 540 FTRGOCVEECRYLQGLPREYVNAHICLPCHPECCQONSGVTFEGEDADQVACHAYKDP 599  
 DB 536 FTRGOCVEECRYLQGLPREYVNAHICLPCHPECCQONSGVTFEGEDADQVACHAYKDP 595  
 QY 600 PCVVARCPGVPKDLSTYMPIMKPEDEGACQCPINCTSHCVLDKGCAPAEORASPLTS 659  
 DB 596 SSCVVARCPGVPKDLSTYMPIMKPEDEGACQCPINCTSHCVLDKGCAPAEORASPLTS 655  
 QY 660 IYSAVVGILLVVVGVFGILLKRRQOKIRKYMRLLOETELVEPLTSGAMPNOQOKR 719  
 DB 656 IYATVVGILLVVVGVFGILLKRRQOKIRKYMRLLOETELVEPLTSGAMPNOQOKR 715  
 QY 720 ILKETELRKVKVLGSAFGTVVYGIWIPGDNKIKVAILKVLRENTSPKANKEILDEAYV 779

DB 716 ILKETELRKVKVLGSAFGTVVYGIWIPGDNKIKVAILKVLRENTSPKANKEILDEAYV 775  
 QY 780 MAGVSPYVSRLLGICLTSTVQLVTLQMPGCLLDHVERNGRLSGODLLNMCQIATKGM 839  
 DB 776 MAGVSPYVSRLLGICLTSTVQLVTLQMPGCLLDHVERNGRLSGODLLNMCQIATKGM 835  
 QY 840 SYLEDRLVHRDLAANVLYKSPNHVKTIDPGLARLLDIDETETHADGCKVPIKMMALLES 899  
 DB 836 SYLEDRLVHRDLAANVLYKSPNHVKTIDPGLARLLDIDETETHADGCKVPIKMMALLES 895  
 QY 900 ILRRRFTHSDVSVYVTLWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYM 959  
 DB 896 ILRRRFTHSDVSVYVTLWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYM 955  
 QY 960 IMVKCMIMSECCPRRELVSERSRMAAPQRFVVIQNEBLGASPLDSTFFYSLLLEDD 1019  
 DB 956 IMVKCMIMSECCPRRELVSERSRMAAPQRFVVIQNEBLGASPLDSTFFYSLLLEDD 1015  
 QY 1020 MGDLYDAEYLVQGFCCPDPAFGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1079  
 DB 1016 MGDLYDAEYLVQGFCCPDPAFGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1075  
 QY 1080 PLAPSEGASDVFDGDLGMAAKGLSLPTHDSPLQRYSEDPVLPSETDGYVAPLTC 1139  
 DB 1076 PLAPSEGASDVFDGDLGMAAKGLSLPTHDSPLQRYSEDPVLPSETDGYVAPLTC 1135  
 QY 1140 SPOPEYVNPDPVRPQPPSPREGPLPAARPAATLEBAKTLSPKNGVYKDVAFAGAVEN 1199  
 DB 1136 SPOPEYVNPDPVRPQPPSPREGPLPAARPAATLEBAKTLSPKNGVYKDVAFAGAVEN 1195  
 QY 1200 PEYLTPOGGAAPQHPPPAFSPAFDNLVYWDOPPERPAPSPSTFFKGPPTAENEYGLDV 1259  
 DB 1196 PEYLTPOGGAAPQHPPPAFSPAFDNLVYWDOPPERPAPSPSTFFKGPPTAENEYGLDV 1255  
 QY 1260 PV 1261  
 DB 1256 PV 1257

RESULT 3  
 ERB2\_MESAU STANDARD; PRT; 1254 AA.  
 AC 060553;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (neu proto-oncogene) (C-erbB-2).  
 GN ERB2 OR NEU.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Nerve;  
 RX MEDLINE=94193007; PubMed=7908275;  
 RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,  
 RA Yamazaki Y., Ishikawa T.;  
 RT Cloning and activation of the Syrian hamster neu proto-oncogene.;  
 RL Gene 140:251-255(1994).  
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIRGULIN (BY SIMILARITY).  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE



CC RESIDUES.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D16295; BAA03801.1; -  
 CC HSSP: P11362; 1FCX.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; kinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TyrKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KM Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1254  
 FT DOMAIN 22 652  
 FT TRANSMEM 653 675  
 FT DOMAIN 676 1254  
 FT DOMAIN 158 358  
 FT DOMAIN 472 644  
 FT DOMAIN 720 987  
 FT NP\_BIND 726 734  
 FT BINDING 753 753  
 FT ACT\_SITE 845 845  
 FT DISULFID 195 204  
 FT DISULFID 199 212  
 FT DISULFID 236 244  
 FT DISULFID 240 252  
 FT DISULFID 255 264  
 FT DISULFID 268 295  
 FT DISULFID 299 311  
 FT DISULFID 315 331  
 FT DISULFID 334 338  
 FT DISULFID 511 520  
 FT DISULFID 515 528  
 FT DISULFID 531 540  
 FT DISULFID 544 560  
 FT DISULFID 563 576  
 FT DISULFID 567 584  
 FT DISULFID 587 596  
 FT DISULFID 600 623  
 FT DISULFID 626 634  
 FT DISULFID 630 642  
 FT MOD\_RES 1139 1139  
 FT MOD\_RES 1247 1247  
 FT CARBOHYD 68 68  
 FT CARBOHYD 125 125  
 FT CARBOHYD 187 187  
 FT CARBOHYD 259 259  
 FT CARBOHYD 530 530  
 FT CARBOHYD 571 571  
 FT CARBOHYD 629 629  
 FT CARBOHYD 658 658  
 FT VARIANT 659 659  
 FT SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 86.5%; Score 5921.5; DB 1; Length 1254;  
 Best Local Similarity 86.7%; Pred. No. 3,6e-310;  
 Matches 1093; Conservative 58; Mismatches 103; Indels 7; Gaps 3;  
 QY 1 MELALCRGGLLALLPPGAASVCTGTDKRLRLPASETHLMDLRHLVGGCVVQGNL 60  
 DB 1 MELAAWCGGLLALLSPASGVCTGTDKRLRLPASETHLMDLRHLVGGCVVQGNL 60  
 QY 61 ELTYPTNALSFLDIOEVGVVLIANQVQVPLQRLRIYRGTOLEFENVLALAVDNG 120  
 DB 61 ELTYPTNALSFLDIOEVGVVLIANQVQVPLQRLRIYRGTOLEFENVLALAVDNR 120  
 QY 121 DPLNNFNFTVSWLRVPRVSAASHLEQLSLTEILKGVLIQRNPOLCYODTILMKDIFH 180  
 DB 121 DPLDNTVATG---RTPE--GLRELQSLSLTEILKGVLIQRNPOLCYODTILMKDIFR 174  
 QY 181 KNNQALTLIDNRSRACHPCSPMCKGSRCKWGBSSDCOSLRTVCAGCCARCKGPLPTD 240  
 DB 175 KNNQALPVDIDNRSRACHPCSPACKCKNHCKWGBSSDCOTLCTGTIAPRAVPARARLPTD 234  
 QY 241 CCEBOCAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPREGRYTFGAS 300  
 DB 235 CCEBOCAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPREGRYTFGAS 294  
 QY 301 CYTACPRYNYLSTDVGSCTLVCPDLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVR 360  
 DB 295 CYTTCPRYNYLSTEVGSCTLVCPDLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLRGAR 354  
 QY 361 AVTSANIOEFACCKKIFGSLAFPLSPFDGDPASNTAPLOPEOLQVETLEITGYLYISA 420  
 DB 355 AVTSANIOEFACCKKIFGSLAFPLSPFDGDPASNTAPLOPEOLQVETLEITGYLYISA 414  
 QY 421 WPDSLPDLVSFONLQVIRIRILHNCAYSLTLOGIGISWGLASLRELGGALIHNTHL 480  
 DB 415 WPDSLHDSVFPONLVRGRVLDGAYSLTLOGIGIRWGLASLRELGGALVLIHNTHL 474  
 QY 481 CFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHMGCGPTCCVNCOSF 540  
 DB 475 CFVHTVPMQDLFRNPHQALLHSGNPSSEECGLKDFACRYLCAHGHCWGSPPTCCVNCOSF 534  
 QY 541 LRQECVEBCRYLQGLPREYVNAHICLPCHPECQPNONGSVTGFGEADOCVACAHYKDP 600  
 DB 535 LRQECVEBCRYLQGLPREYVNAHICLPCHPECQPNONGSVTGFGEADOCVACAHYKDP 594  
 QY 601 FCVARCPGVKXKDLSTYMPIMKRPDEEGACQPERINCTHSVDLDDKGCAGAEQASLTSI 660  
 DB 595 FCVARCPGVKXKDLSTYMPIMKRPDEEGACQPERINCTHSVDLDDKGCAGAEQASLTSI 654  
 QY 661 VSAVVGILLVVLGVVFGILIKRPOOKIRKYMTRLOETELVEPLTPSGAMPNOQMRI 720  
 DB 655 IATVVGILLVVLGVVFGILIKRPOOKIRKYMTRLOETELVEPLTPSGAMPNOQMRI 714  
 QY 721 LKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYVM 780  
 DB 715 LKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYVM 774  
 QY 781 AGVSPYYSRLIGLITSTVQLVTLMPYGCLLDHYREKRGVLSODLLNMQOIAKGS 840  
 DB 775 AGVSPYYSRLIGLITSTVQLVTLMPYGCLLDHYREKRGVLSODLLNMQOIAKGS 834  
 QY 841 YLEVDRLVHRDLAARNVLYKSPNHKITDFGLARLLIDETEVHADGGKVPKKNMALSEI 900  
 DB 835 YLEVDRLVHRDLAARNVLYKSPNHKITDFGLARLLIDETEVHADGGKVPKKNMALSEI 894  
 QY 901 LRRRFTQSDVWSGVTVWEIMTFGAKPYDGIIPAREIDLEKGRLLPQPICTIDVYMI 960  
 DB 895 LRRRFTQSDVWSGVTVWEIMTFGAKPYDGIIPAREIDLEKGRLLPQPICTIDVYMI 954  
 QY 961 MVKCMIDSECRPRRRELVSFESRVARDPORFVVIQNDLDGPASLDSTFRSLLEDDM 1020  
 DB 955 MVKCMIDSECRPRRRELVSFESRVARDPORFVVIQNDLDGPASLDSTFRSLLEDDM 1014

QY 1021 GDLVDAEEYLVPOQGFCEPDPAAGAGVWHRRSSSTRSGGDLTLGLPESEEAPRSP 1080  
 DB 1015 GDLVDAEEYLVPOQGFCEPDPAAGAGSTARRRRSSSTRSGGDLTLGLMEPSGEPPRSP 1074  
 QY 1081 LAPSEGAGDVDPDGLGMAAGLQSLPHHDSPLQRYSEDPVPLPSTTDGVVALTGS 1140  
 DB 1075 LAPSEGAGDVDPDGLGMAAGLQSLPHHDSPLQRYSEDPVPLPSTTDGVVALTGS 1134  
 QY 1141 POPEYVNOVDVPRPQPPSPREGPLPAARPAATLERAKTSPGKNGVVKDVFARFGAVENP 1200  
 DB 1135 POPEYVNOVDVPRPQPPSPREGPLPAARPAATLERAKTSPGKNGVVKDVFARFGAVENP 1194  
 QY 1201 EYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPERGAPEPTFKGPTAENPEYLGADV 1260  
 DB 1195 EYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPERGAPEPTFKGPTAENPEYLGADV 1253  
 QY 1261 V 1261  
 DB 1254 V 1254

RESULT 4  
 EGFR\_HUMAN STANDARD; PRT; 1210 AA.  
 ID EGFR\_HUMAN  
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;  
 AC P00588; Q9B2S2; Q9B2C9; Q9GZC9; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 protein-tyrosine kinase ErbB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Charyocota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Mayer E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";  
 RL Nature 309:418-425(1984).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=95382957; PubMed=7654366;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN (3)  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97078686; PubMed=8918811;  
 RA Reiter J.L., Mailhe N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN (4)  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Garfili J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN (5)  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX TISSUE=Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Scheel Sinclair C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Mailhe N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Genomics 71:1-20(2001).  
 RN (6)  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Scheel C.M.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
 RA Mailhe N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN (7)  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krutiger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells.";  
 RL Science 224:843-848(1984).  
 RN (8)  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAe overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810(1984).  
 RN (9)  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN (10)  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";  
 RL Oncogene Res. 1:375-396(1987).  
 RN (11)  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis.";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN (12)  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN (13)  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN (14)  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Wroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and supercoiled DNA." ;  
 RL Nature 309:270-273(1984).  
 RN (151)  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honneger A.M.,  
 RA Howk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HHR23/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor." ;  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN (16)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts." ;  
 RL Growth Factors 13:121-132(1996).  
 RN (17)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668;  
 RA Sato C., Kim U.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor." ;  
 RL J. Biochem. 127:65-72(2000).  
 RN (18)  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Otake M., Inagaki F., Lax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor." ;  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN (19)  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens." ;  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC - FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC - FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC - ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4, are produced by  
 CC alternative splicing.  
 CC - TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC - MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, X00588; CAA25240.1; -  
 CC DR EMBL, U95089; AAB53063.1; -  
 CC EMBL, U48722; AAC50802.1; -

DR EMBL, U48723; AAC50804.1; -  
 DR EMBL, U48724; AAC50796.1; -  
 DR EMBL, U48725; AAC50797.1; -  
 DR EMBL, U48726; AAC50798.1; -

Query Match 45.8%; Score 3137; DB 1; Length 1210;  
 Best Local Similarity 49.3%; Pred. No. 6,6e-161;  
 Matches 627; Conservative 178; Mismatches 352; Indels 116; Gaps 22;

11 LLLALLPQGA--STVGCTGDMKLLPASPETHLDMRLYGGCOVVOGNLELTYPTN 68  
 14 LLLALCPASRLAEKKVCCGTSNKLTLQGLFDFHPLSLQMFNNCEVVLGNLEITYVRN 73  
 69 ASLSFLDIOEYGVYLIANQVYPLRLIRVGTQQLFEDYVALAVDNGPLNFFN 128  
 74 YDSFLKTIQEVAGVYLIALNTERLPLENLQIRRMYYENYALAVLSNYD----- 126  
 129 FTVSWFLRVKVSASHLEOL--RSLTELLKGVLIQNPOLCYODTTLMKDFHKNOLA 186  
 127 -----ANKTGLKELPMRNLOEILHGAVRFSNNPALCNVESIQMRDIVSDPLSN 175  
 187 LTLIDTNRBRACHPCSPMKSGRCWGESSEDDQSLLTRYCAGGCA-RCKGPLPTDCHEQ 245  
 176 MSMDFFQNLGSCQKCDPCSPNCSWGAGBENCQKLTKIICAOQCSGRCRKSPSDCHNQ 235  
 246 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPFGRTFGASCYTAC 305  
 236 CAAGCTGPRESDCLVCRKRDEATCKDTGCPPLMLYPTTYQMDVNEBKGTSFATCYKCC 295  
 306 PNYVISTDVGCTTLCVPLHNOEVTAEQTORCEKSKPCARVCYGLMEHLREAVTSA 365  
 296 PRNYVVTDHGSCVRACGADSYEM-EDGVKCKCKCGCRKVCNGIGIFKSLISINAT 354  
 366 NIQEFAGCKKIGSLAFLPESPDGPDASTAPLQPOLQVETLEITGLYLSANPDSL 425  
 355 NIKHFKNCTSIGDHLPLVAFRGDSFTHTPLDPELDLTKVKEITGFLLLQAMPENR 414  
 426 PDLVSFQNTQVIRGRILIHNGAVSLTQGGISWLGSRSLRELGGSLALLHNNHLCFVHT 485  
 415 TDLHAFENLEIRGRKQKQGFSLAVSLNTSLGRLSKELISDGVIIISGNKULCYANT 474  
 486 VPMDOLEFRNPQALLHTANRPEDECVGEGLACHQLCARGHGMGPPTOCVNCQSOLRGOE 545  
 475 IMWKLFSGISGQTKLIISNRGENSCATQOVCHALCSPGCMGPEPRDVCSCNVSGRE 534  
 546 CYBECVYLOGLPREYVNAHCLPCHPQCQPNQSVTCFREPADQVCAAHYKPPCVAR 605  
 535 CVDKCNLLEGEPRFENSECICQHCPCLPQAMNITCTGRGPNCTIQCAHYIDGPHCVKT 594  
 606 CPSPGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGPAPGQASPLTSISAVY 665  
 595 CQAGVWGENNTL-VMKRYADAGVCHLCHENCTYCTGPGLEGCPITNGPKIP--STATGV 651  
 666 G---LLVVLGVFVGLILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQOMRILK 722  
 652 GALLLLLVVALGIG--LFMRBRHIVRKTLRLRLOERLVEPLTPSGEAPNALRILK 708  
 723 ETELKRVKVLGSGAGSTYKGIWIPDGENVKIPVAITKURENTSPRANKETIDEAYVAG 782  
 709 ETEFKIKIVLGSGAFITVKKGLMPEGEKVIIPVAIKELREATSPRANKETIDEAYVAG 768  
 783 VGSPPVSRLLGLCTNSTVQLVNTQMLPVGCLLDHVRENRRLSGODLLNNCMQIAKMSYL 842  
 769 VNPVHVCRLGLCLTSTVQLTQLMHPGCLDVTREHKONIGSYLLNMCVQIAKMANYL 828  
 843 EDVRLVHRDLAARVLYKSPNHVKITDPGLIARLLDIDETEHADGGKVPDKMVALESIR 902  
 829 EDRLRVHRDLAARVLYKTPQHVKITDPGLIARLLDIDETEHADGGKVPDKMVALESIR 888  
 903 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKBERLPQPICTIDVYMW 962  
 889 RIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISISLEKGERLPQPICTIDVYMW 948

```

QY 963 KCMWIDSECRPRELWSEFSRMARDPQRFVIO-NEDLGAPSLDSTFYRSILEDMDG 1021
DB 949 KCMWIDADSRKRFRELIERFSKMAROPQRLVIGQDBRMHLPSFTSDNFIRALMDEEDMD 1008
QY 1022 DLVDAEYLVPOGCFPCDPAPGAGCMVHRHRSSTRSGGDLTLCLEPSEEBAPSP 1081
DB 1009 DVDVDADEYLVPOGCFPCDPAPGAGCMVHRHRSSTRSGGDLTLCLEPSEEBAPSP 1034
QY 1082 APESGSDVDFDDDLGMAKGLQSLPTHPSPLORESEDPVLPSET--DGVAVLTC 1139
DB 1035 LSLSLATSN--NSTVACIDNGLGSCPIKEDSFQRTSSDPTGALTEDSIDDTFL----- 1087
QY 1140 SPOPEYVNPQDVPRPQPPSPREGPLPAAPAGATLERAKTLPGRNGVYKDVFAFGAVEN 1199
DB 1088 -PVEYINQ-SVPRKPAQSVQNPVYHNQPLNP-----APSRDPHYQD--PSTAVGN 1135
QY 1200 PEVL-TPGGAAAPQHPPPAPFPAFDNLTYWDQ-----DP-----PERGAPST 1242
DB 1136 PEYLVNTVQ-----PTCVNSTFDSPAHMAQKSHQISLDNPDYQODEFPPEAKENG 1186
QY 1243 FKCTPAENPEYL 1255
DB 1187 FKGS-THENAYL 1198

RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
ID EGFR_MOUSE
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yavon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

```

```

RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@sib.ch).
CC
DR EMBL: X78987; CA55587.1; -
DR EMBL: U03425; AAA1789.1; -
DR EMBL: X59698; CAA4221.1; -
DR EMBL: L06864; AAA53029.1; -
DR EMBL: Z12608; CAA78249.1; -
DR HSSP: P11362; 1FGK.
DR WGD; MG1:95294; Egfr.
DR Interpro: IPR000494; EGFR_L domain.
DR Interpro: IPR000719; Euk_pkinase.
DR Interpro: IPR002174; Furin-like.
DR Interpro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L domain; 2.
DR Prodom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fu; 3.
DR SMART: SM00219; TyrKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 1210
FT DOMAIN 25 647
FT TRANSMEM 648 670
FT DOMAIN 671 1210
FT REPEAT 75 300
FT REPEAT 390 600
FT DOMAIN 1028 1071
FT DOMAIN 714 981
FT NP BIND 720 728
FT BINDING 747 747
FT ACT SITE 839 839
FT DISULFID 190 199
FT DISULFID 194 207
FT DISULFID 215 223
FT DISULFID 219 231
FT DISULFID 232 240
FT DISULFID 236 248
FT DISULFID 251 260
FT DISULFID 264 291
FT DISULFID 295 307
FT DISULFID 311 326
FT DISULFID 329 333

```

```
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46D2D2F5 CRC64;

Query Match 45.4%; Score 3109; DB 1; Length 1210;
Best Local Similarity 49.1%; Pred. No. 2,1e-159;
Matches 629; Conservative 171; Mismatches 360; Indels 120; Gaps 24;

Qy 11 LLLALLPFGAA--STQVCTGTDMLRLPASPTLHMLRLHLYQGCQVVOGNETLYLPTN 68
Db 14 LLLTLLCAAGALAEKKVCOGTSNRLTQGFEDHFLSLQRMVNCCEVVLGNLEITYVQRN 73
Qy 69 ASLSFLDIOIEVOGVYLLAHNOVROVRLRIVRGTOLEFEDVYALAVLDNDPDLNPRN 128
Db 74 YDLSFLKTIQEVAGVYLLAINTVERIPLENDOIIGNALYENTVYLLALISNG----- 126
Qy 129 FTVSFWLRVPRKVSASHLEQL--RSLTLEILKGGVLIQRNPOLCYODTILMKDI----FKN 182
Db 127 -----TNRGLRELPMRNLQELILGAVFRSNPILCMOTLIQMRDLVQNVFMSN 175
Qy 183 NQLALTLIDTRSRACHCSPMKCSRCESESECCSLTRTVCAAGCA-RCKGPLPTDC 241
Db 176 MSMDL---QSHPSGCPKCDPSCPNGSCMGGEENCOKLTKIICAQOCSHRCGRSPDC 231
Qy 242 CHEQCAAGCTGPKHSDDCLAGLHFNHSGICELHCPALVYNYDTFESMNPREGRYFGASC 301
Db 232 CHNCAAGCTGPRSDCLVCCQKFODEATCKOTCPRLMYNPTTYOMDVNBEKYSFGATC 291
Qy 302 VTACPRYNYLSTDVGSCTLVCPRLHNOEVTAEADGTORCEKSKPCARVACVGLGMEHLREYRA 361
Db 292 VKKCPRYNYVVDHSGCVNACGPDYEV-EDDGIRKCKKCDGCRVACVGLGMEHLREYRA 350
Qy 362 VTSANIQEFACGCKKIFGSLALUPSEFDDPASNATPRLPEOLQVETLEITGYLYISAM 421
Db 351 INANINIKFKYCTAISGDLHLIPVAFKGDSPFRTPRLPRELEIKTYKEITGFLLIQAW 410
Qy 422 PDSLPLDLSVFONLQVIRGRIILHNGAYSLTLOGLIGMTLSRLREISGLALIHNTILC 481
Db 411 PDNMTDLHAFENLEIRGTRKOHQGFSLAVVGLNTISGLNSLKSISGDVYIISGNRLVC 470
Qy 482 FVHTVPMVDQLEFRNPHQALLHTANREDECEVBGLACHQLCARHGCWGPPTQCVNCSOFL 541
Db 471 YANTINWKKLFGTPNQKTKIMNNAEKDCKAVNHVCNPLCSSEGCGWGPPTQCVNCSOFL 530
Qy 542 RGGCEVEBCRVLOGLRREYVNAARHLCPEHRECQFONGSVTCRGPEADCCVACAHYKDPDF 601
```

```
Db 531 RRECEKKNILGEBREFEVENSECICQCHPECLPQAMNITTCGRGDNOCICQAHYIDGPH 590
Qy 602 CVARCPGKVPDLSYMPIWKFPEDEGACCPCPINCHSCVDLDDKCCPAEQRASPLTSIV 661
Db 591 CKTTPAGIMGSENNL-VKTYADANNVCHLCHACTYGCAGGELQCEWPPSPCKLPSIA 649
Qy 662 SAIVGILLVNVYGVVFGI-LIKRQOKIRKTYMRRLLOTELEVERLTPSGAMPNOQMRI 720
Db 650 TQIVGGLFLIV-VALGIGLFMRRIYVKRTLRRLQRELEVERLTPSGEARNQHLRI 708
Qy 721 LKETELRKVKYLGSGAFGVYVYGIWIPDGENYKIPAIYVLRBNTSPKANKELIDAYVM 780
Db 709 LKETEFKIKYVLSGSAFGVYVYKGLWIPSEGEKYIPVAIKELREATSPKANKELIDAYVM 768
Qy 781 AGVSPVYSRLIGICLSTVOLTOLMPYGCILLDHYREKRGUSGDDLNMWCOIKGMS 840
Db 769 ASVDNPHVCRLLGICLSTVOLTOLMPYGCILLDHYREKRGUSGDDLNMWCOIKGMS 828
Qy 841 YLEDVRLVHRDLAANNVYKSPNHVYKTDPGFLARLIDIDETEHADGKVPKIMALESI 900
Db 829 YLEDRLVHRDLAANNVYKTPQHYKIDFGAKLILGAEKETHAEGKVPKIMALESI 888
Qy 901 LRRRTTQSDVWSYGVYTWELMTFGAKPYDGIIPAREIPDLERGERLPQPICTIDVYMI 960
Db 889 LHRITTHOSDVSYGVYTWELMTFGSKPYDGIIPASDISILEKGERLPQPICTIDVYMI 948
Qy 961 MWKCMIMDSECPRERELVESESRMARDPQRVVVQ-NEDDLGPASPLDSTFRSLLEDD 1019
Db 949 MWKCMIMDSDSPKRELLEFSKMARDPQRYLVIGQDHRMLPSPDTSNFRLALDEED 1008
Qy 1020 MGDLDVAEYLVYPOQGFPCPDPAAGGVNHRHSSSTRSGGDLTLLEPSEEEAPS 1079
Db 1009 MEDVDADAEYLIPQGGF-----NSPST-----SKT 1034
Qy 1080 PLAPSEAGSDVFDGLGKAAGLQSLPTHDPSPLOYSEDPYVLPSET--DGVAAPL 1137
Db 1035 PLSSLSATSN-----NSTVACINRNGSCVKEDAFQRYSSDPTGAVTEINDIDATL--- 1087
Qy 1138 TCSPPQEVYNODVAPPOPPSPREGPLPAARPAATLERAKTISPQKNGVYKDVAFAGAV 1197
Db 1088 ---PVPEYVNO-SVPRKRPAGVQNPVYHNPQLHP-----APGDHLHYQN--PHSNAY 1133
Qy 1198 ENPEYL-TPQGAQAQPPHPAPFAPDLVYWD-----DP-----PERGAP 1240
Db 1134 GNPEYLNTAQ-----PTCLSSGFSNPSALMIQKSHQMSLDNPDYQODFFPKTKPR 1184
Qy 1241 STFKGTPAENPEYGLDVP 1260
Db 1185 GIFKG-PTAENAEYLRAVP 1203

RESULT 6
ERB4_HUMAN STANDARD; PRT; 1308 AA.
ID ERB4_HUMAN
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-tye cell surface receptor HER4).
GN ERB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma.
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.,
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
```

FT	NP_BIND	724	732	ATP (BY SIMILARITY).
FT	BINDING	751	751	ATP (BY SIMILARITY).
FT	ACT_SITE	843	843	BY SIMILARITY.
FT	DISULFID	189	197	BY SIMILARITY.
FT	DISULFID	193	205	BY SIMILARITY.
FT	DISULFID	213	221	BY SIMILARITY.
FT	DISULFID	217	229	BY SIMILARITY.
FT	DISULFID	230	238	BY SIMILARITY.
FT	DISULFID	234	246	BY SIMILARITY.
FT	DISULFID	249	258	BY SIMILARITY.
FT	DISULFID	262	289	BY SIMILARITY.
FT	DISULFID	293	304	BY SIMILARITY.
FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	VARSPLIC	626	648	NGPSHDCIYYPMCHSPLPOHA -> IGSSTDICIGLMD (IN ISOFORM JM-B).
SQ	SEQUENCE	1308 AA;	146807 MW;	5E4AE8095D08761 CRC64;
Query Match		43.6%;	Score 2986.5;	DB 1; Length 1308;
Beat Local Similarity		45.2%;	Pred. No. 8.5e-153;	
Matches	612;	Conservative	185;	Mismatches 374; Indels 183; Gaps 29
OY	9 WGLLIALLIPGAA----	STOVCTGTDMKRLRPASPETHLDMLRHLYGCCOYVOCNLELTYY	64	
Db	8 WWWSLTVAAGTVGPDSQSVCACTENKLSSLDIEQQYALARKYENCVEVMGNLBITS	67		
OY	65 LPTNASISFLDDIOEVGVYLIAHVNOVQVLQQLRIYRGQLPEDNYALAVLNDGDPLN	124		
Db	68 IEHRNDSEFLSVREVTGYLVVALNQFYLPLENIRLRIGKLVLEDYRALAIFLNYRKDG	127		
OY	125 NFNNFTVSFWLRVPKVASHLEQ--LSLSYTEILKGSVLIQRNPOLCYODTLIMKDIFKHN	182		
Db	128 NFG-----LQELGKNLTELINGVVVDNMKFLCVAADTIHMQDIYANP	170		
OY	183 NQALLTLIDITRSRACHPCSPMCKSGSRWESSSEDCCSLRTTVACAGC-ARCKGPLPTDC	241		
Db	171 WPSNLTLIVSTGSSGCCGRCHKSCTG-RWMGETEHNCQTLLRTTVACEDCCDGCRCYPVYSDC	229		
OY	242 CHECGAAGCTCPKPSDCLACLFHNSGICELHCPALVLYNPDTESMNPBGRYTFFGASC	301		
Db	230 CHRECAAGCGSPKDTDPACAKNFNDSAGCACTQCQTGYFNNTTTQULEHNFAKTYYGAFc	289		
OY	302 VTACPYNVYSTDVSSCTLVCPLAHNQEYTAEDGTORCEKSKPCARVCYGLMEHLREVRA	361		
Db	290 VKCKPNHFV-VDSSSCVARACPPSSIKMEV-EENGIMCKRPTCIDCPKACDIGIGTSIMSNOT	347		
OY	362 VTSANIDEFAGCKKIPOSIALFPESPFDGDPASNAPRLQPBPLOVVFETLLEETIGTYLISAW	421		

```

Db 348 VDSNDKFNCTKINGNLIFLVYGIHGPVNAIEAIDPEKLNVPRTVEITGFLNIQSW 407
Qy 422 PPSLDLSFQNLQVIRGILHNGASVLTQGLGIMLGRSIRELGSGLALIHNTLHC 461
Db 408 PNMMDVFSVSNLVYIGGRVLSGLSLTLKQOGLSTLQFOSTLKEISAGNIYITDINSJC 467
Qy 482 FHTYPMDOLEFNPPOHALLHTANREDECEGGLACHOUCARHCMPGPGTQCVNCSQFL 541
Db 468 YHTITMTLTFSTINORIVIRDRKAKENCTABGMVCHLSSDGCMPGPGDQCLSCRPS 527
Qy 542 RQOECEVEECLVQLGPREVYNAHCLPCHRECP-ONGSVTCGPEPADOCVACAHKDP 600
Db 528 RGRICIESCNLYDGEFREFENSICVECDPOCEKMGDLTTCGPEPDNCTKSHKDP 587
Qy 601 PCVACPSGVKPDLSMPIMKPEDEGAQPCPCINCTHSCVDLDKGC-----P 649
Db 568 NCVKCPDGLQGANSE-IFKYADPRECHPCHPNCTOCNGPTSHDCTIYPTWGHTLP 645
Qy 650 AQRASPLTSIVSAVV-GILLVVLGVNFGILIKRQOKIRKTKMRLLOETELVPLRP 708
Db 646 QHAR-TPL--LAAGVIGLFLIYVGLTAVVRRASIK-KRALRFL-ETELVPLRP 700
Qy 709 SGAMPNOAMRLKETELKRVKVLGSGAFVYKGIWIDGENVKIPVAIKVLENTSPK 768
Db 701 SGTARNQQLRLKETELKRVKVLGSGAFVYKGIWIDGENVKIPVAIKVLENTSPK 760
Qy 769 ANKELIDEVYVAGSPVSVRLGICLTSTVQLVQLMPYGCGLDHPENRGLSGDL 828
Db 761 ANVEFMDEALIMASNDHPLVRLGLVCLSPITQVLQMLPHGLLLEYVHEHMDNLSQL 820
Qy 829 LNMCOIAKMSYLEEDVRLVHRDLAARNVLYVSPNIVK1TDPGLARLLIDETEVADGG 888
Db 821 LNMCOIAKMSYLEEDVRLVHRDLAARNVLYVSPNIVK1TDPGLARLLIDETEVADGG 880
Qy 889 KVPIMKMALESILARRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 948
Db 881 KVPIMKMALESILARRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 940
Qy 949 OPTICTIVYMMVCMVIDSCRRFRFELVSEFSRMAPDPORFVYIQNED-LGPASPLD 1007
Db 941 OPTICTIVYMMVCMVIDSCRRFRFELVSEFSRMAPDPORFVYIQNED-LGPASPLD 1000
Qy 1008 STFYRSLLEDMDGLVDAEEYLVPOQGFPCDPAPAGAGMHHHRSSTSGGDLTL 1067
Db 1001 SKFPNLDEEDLEMDMAEELVLP-QAFNIRP-----ITSRRIDSNSS-----EI 1048
Qy 1068 GLEPSEEBAPRS-----PLAP-SEGAGSDVPFGDLGM 1098
Db 1049 GHSPPPAYTPMSGNGFYVRDGGFAAEGVSVRYAPRTSTIPAPVAGATATIFDSCN 1108
Qy 1099 GAAKGLQSLPTHDPFLQRYSDPVPPLPS-----ETDGYVAPLTCSPQEVYVNOQDV 1151
Db 1109 GTLRPVAPVHVOEDSTQRYSDPTVPAPERSPRGELDEEGYMTMRDKPEQEYLVNPEV- 1167
Qy 1152 RPOPSRPEGPLPARPAGATLERAKTISPCKNGVAVKOVAFAGCAVENEYVITPGGAAP 1211
Db 1168 -----ENPFVSR-----KNGDLQ-----ALNDENYNASNG--- 1194
Qy 1212 QPHPPA-----FSPAFDNLVYWDQDPERG 1237
Db 1195 ---PPKADEYVNEFLYNTFANTLGAKAEYLNKNILSMPEKAKKAFDNDPVYNNHSLPRS 1251
Qy 1238 A--PPSTKGTPT-----AENPEYL 1255
Db 1252 TLQHPDYLQEYSTKYFYKNGRIRPVAENPEYL 1285

```

```

RESULT 7
ERR4 RAT STANDARD: PRT: 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Babiaga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulin promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Fromme P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
RN [4]
RP FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN. NRG-
RP 2. NRG-3. HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
RP NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
RP NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF041838; AAD08899.1; -.
CC EMBL: U52531; AAC53051.1; -.
CC HSSP: P11362; 1FGK.
CC InterPro: IPR000494; EGFR_L domain.
CC InterPro: IPR000719; Euk_Pkinase.
CC InterPro: IPR002174; Eukin-like.
CC InterPro: IPR001245; Tyr_Pkinase.
CC InterPro: IPR004019; YLP_motif.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF00069; Pkinase; 1.
CC Pfam: PF01030; Recep_L_domain; 2.
CC Pfam: PF02757; YLP_2.
CC PRINTS: PR00109; TYRKINASE.
CC ProDom: PD000001; Euk_Pkinase; 1.

```



```

DR SMART; SM00261; FU; 4.
DR PROSITE; PS00109; TYRKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Trastembiane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT CHAIN 1 25
FT SIGNAL 1 25
FT DOMAIN 26 1308
FT TRANSMEM 652 675
FT DOMAIN 676 1308
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 718 985
FT NP_BIND 724 732
FT BINDING 751 751
FT ACT_SITE 843 843
FT DISULFID 189 197
FT DISULFID 193 205
FT DISULFID 213 221
FT DISULFID 217 229
FT DISULFID 230 238
FT DISULFID 234 246
FT DISULFID 249 258
FT DISULFID 262 289
FT DISULFID 293 304
FT DISULFID 308 323
FT DISULFID 326 330
FT DISULFID 503 512
FT DISULFID 507 520
FT DISULFID 523 532
FT DISULFID 536 552
FT DISULFID 555 569
FT DISULFID 559 577
FT DISULFID 580 589
FT DISULFID 593 614
FT DISULFID 617 625
FT DISULFID 621 633
FT MOD_RES 1162 1162
FT MOD_RES 1188 1188
FT MOD_RES 1258 1258
FT MOD_RES 1284 1284
FT CARBOHYD 138 138
FT CARBOHYD 174 174
FT CARBOHYD 253 253
FT CARBOHYD 358 358
FT CARBOHYD 410 410
FT CARBOHYD 473 473
FT CARBOHYD 495 495
FT CARBOHYD 548 548
FT CARBOHYD 576 576
FT CARBOHYD 620 620
FT CONFLICT 1062 1062
FT CONFLICT 1080 1082
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 43.3%; Score 2967; DB 1; Length 1308;
Best Local Similarity 45.0%; Pred. No. 9.4e-152;
Matches 610; Conservative 192; Mismatches 383; Indels 170; Gaps 29;

```

```

QY 176 KDIFFHKNQALATLIDITNRSRACHPCSPKCGSRGCESESDCSLTRTVACGC-ARCK 234
DB 164 QDIWRPSPNMTLVSTIGSSGGRCHKSTG-RCKGPTFNHCOTILRTVACACDGRCV 222
QY 235 GPLPTDCHEQCAAGTGPRHSDCLALFNHNSGICELHCPALVTYNTDTFESMPDEGR 294
DB 223 GPVNSDCRCHREACGSGSPDTCFACMFNFNDGACVTCOPQFFVFNPTTFQLEHNFNK 282
QY 295 YTFGASCVTACPTNYLSTDVGSCTLVCPRLNQVTLADGTQREKSKSPARCVYIGME 354
DB 283 YTYGACVCKCPNPFV-VSSSCVRACPSKMEV-BENGIKMKPCPTDLCPKACDGTGTG 340
QY 355 HIREVAVTSANIOEAGCKIIFGSLAFLPESEFGPASNTABLOPELOVFELEETIG 414
DB 341 SLMSAQTVSSNDIKINCKINGNLIFLVTHGHPYNAIDAIDEKLVNFTVREITG 400
QY 415 YLISAMPDLSPLSVFQNLQVIRGRILHNGAVSLTLQGLGSMGLRSIRELGSGLAL 474
DB 401 FLNIQTWPPNMTDFSVFSLNLTIGRHVLSGLLILKQGGITSLQFOSLKEISAGN1YI 460
QY 475 HNHTHLCFHTVPMWDLFENPHQALLHTANRPDECEVGEGLACHOLCARGHCGRPTOC 534
DB 461 TDNSNLCTYHTINMTTLFSTVQRIIVRDRRARENCTAEGMVCNHLCSNDGCGPGDQC 520
QY 535 VNSOPLRQOECVEECRVLOGLPREYVNAHCLPCHPECOF-ONGSVTCFGRPADOCVAC 593
DB 521 LSGRRSRGKICIESCNLIDGFEFEENGISICVCSQCEKMDGLITCGRPDNTKC 580
QY 594 AHYKDPFVCARPCSGVKPDLSPYMPKPEDEBGAOCPEINCTHSCVDLDDKGC----- 648
DB 581 SHKQDQNCVEKCPVLOGANSE--IFKADQRECHPCRPCTGCGNGPTSHDCIYYPW 638
QY 649 -----PAORASPLTISYSAV-GILLVVLGVGILIKRQOKIRKTYRRLLDOTE 701
DB 639 TGHSTLPQAR-TPL--IAAGVIGFLIVIMLTFAVYVRRKSIK-KKALRRFL-ETG 693
QY 702 LVEPLTPSGAMPQAOAMRIKETELRKVKVLSGAGFYKGIWIPDGENVKIPVAIKVL 761
DB 694 LVEPLTPSGARPQAOALRIKETELRKVKVLSGAGFYKGIWIPDGENVKIPVAIKVL 753
QY 762 RENTSPKANKEILDEAYVNAVGSFVYSRLGICLTSTVQLVTLQMPYCGLLDHVRENG 821
DB 754 NETTGKAVNEFDEALIMASVDHPHVLRLVGLCSPTQLVTLQMPHGCLEEVYEHKD 813
QY 822 RUGSODLLMWCQIAGMSYLEVRLVHNDLARNLVYSPNVIKTTDFGLALLDIDET 881
DB 814 NGSQILLMWCQIAGMVLLEERLVHNDLARNLVYSPNVIKTTDFGLALLDIDET 873
QY 882 EYHADGKVPYKMMALLESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDL 941
DB 874 EYHADGKVPYKMMALLESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDL 933
QY 942 EKEGERLPPICTIDVYIMVYKCMIDSECRPRFRLVSEFSSMAEDPQRFVYIONED-L 1000
DB 934 EKEGERLPPICTIDVYIMVYKCMIDSECRPRFRLVSEFSSMAEDPQRFVYIONED-L 993
QY 1001 GPASPLDSTFYSLLEDDDMGDLVDAEYLVPQGGFCPP----- 1041
DB 994 KLPSPDSDKPFQNLDEBLEDMDMAEYLVP-QANFIPPIYTSRTRIDSNSEIGHSP 1052
QY 1042 ----ARGAGMVIHRRSSSTRSGGDLTLGLPESEEARSPRLASSEGASVFPQDGLG 1097
DB 1053 PRAYTMSGSAFYDQGGFATQGG---MMPRYATITSTIREAVVA--QGATAMFDDSCC 1107
QY 1098 MGAAGLQSLPTHPSPRLQRYSEDPVPLPS-----ETDGYAVLTCSPQRYVNOPD 1150
DB 1108 NGTLRKPVVPHVQEDSTQRYSDPTVFAPERVPRAELDEBEGYGTPMHDKRQEVILNPE 1167
QY 1151 VRRQPSRREGRPLARAPAGATLERAKTSLSPGNGVYKQVFAFGAVENPREYLTPOGGA 1210
DB 1168 -----ENPVSRR-----KNQDLO-----ALDNPVYSASG-- 1194
QY 1211 PQPHPPA-----FSPAFDNLVYVDQPPRR 1236

```



```

Db      1195 ----PPKADDEVNEPLYNTFTNALGNAEYMKNSILSVPEKAKKAFDNDPWNHSLPPR 1250
Oy      1237 GA--PPSTFKGTPT-----AENPEYL 1255
Db      1251 STLGHDPVLYQGYSTKYFKYKNGRIRIRVAENPEYL 1285

RESULT 8
XMRK_XIPMA STANDARD; PRT; 1167 AA.
AC      P1338;
DT      01-JAN-1990 (rel. 13. Created)
DT      16-OCT-2001 (rel. 40. Last sequence update)
DE      15-JUN-2002 (rel. 41. Last annotation update)
DE      Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN      XMRK OR TU.
OS      Xiphophorus maculatus (Southern platyfish).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC      Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC      Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX      NCBI_TaxID=8083;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=90015140; PubMed=2797166;
RA      Witbrodt J., Adam D., Malitschek B., Mauelel W., Raulf F.,
RA      Telling A., Robertson S.M., Schartl M.;
RT      "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT      inducing Tu locus in Xiphophorus.",
RL      Nature 341:415-421 (1998).
RN      (2)
RP      REVISION TO 515.
RA      Schartl M.;
RL      Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC      - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC      - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      - SUBCELLULAR LOCATION: Type I membrane protein.
CC      - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC      - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; X16891; CAA34770.2; -
DR      PIR; S06142; S06142.
DR      HSSP; P11362; 1FGK.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_Pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR002290; Ser_thr_Pkinase.
DR      InterPro; IPR01245; Tyr_Pkinase.
DR      Pfam; PF00069; pkinase; 1.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00261; FU; 5.
DR      SMART; SM00220; S_TKC; 1.
DR      SMART; SM00219; TYRK; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR      Trnsmembran; Glycoprotein; Receptor; Signal; Transferase;
DR      Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT      SIGNAL 1 25
FT      CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

```

```

FT      DOMAIN 26 642 KINASE.
FT      TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
FT      DOMAIN 666 1167 POTENTIAL.
FT      DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
FT      NP_BIND 716 724 PROTEIN KINASE.
FT      BINDING 743 743 ATP (BY SIMILARITY).
FT      ACT_SITE 835 835 ATP (BY SIMILARITY).
FT      DISULFID 195 204 BY SIMILARITY.
FT      DISULFID 199 212 BY SIMILARITY.
FT      DISULFID 220 228 BY SIMILARITY.
FT      DISULFID 224 236 BY SIMILARITY.
FT      DISULFID 237 245 BY SIMILARITY.
FT      DISULFID 241 253 BY SIMILARITY.
FT      DISULFID 256 265 BY SIMILARITY.
FT      DISULFID 269 296 BY SIMILARITY.
FT      DISULFID 300 311 BY SIMILARITY.
FT      DISULFID 315 330 BY SIMILARITY.
FT      DISULFID 333 337 BY SIMILARITY.
FT      DISULFID 504 513 BY SIMILARITY.
FT      DISULFID 508 521 BY SIMILARITY.
FT      DISULFID 524 533 BY SIMILARITY.
FT      DISULFID 537 553 BY SIMILARITY.
FT      DISULFID 556 569 BY SIMILARITY.
FT      DISULFID 560 577 BY SIMILARITY.
FT      DISULFID 593 615 BY SIMILARITY.
FT      DISULFID 618 626 BY SIMILARITY.
FT      DISULFID 622 634 BY SIMILARITY.
FT      CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 1167 AA; 129934 MW; 4793B0749DC1D55A CRC64;

Query Match 39.4%; Score 2695.5; DB 1; Length 1167;
Best Local Similarity 45.1%; Pred. No. 3e-137;
Matches 576; Conservative 163; Mismatches 387; Indels 151; Gaps 28;

Oy      4 AALCGWGLLLALPPGAAS-----OVCGTGDMKRLRPSPEHLDMLRLYGGCOVQON 59
Db      8 AALLD--LLLVLSIRCCSTDPDRKVCQSTSNQMTM---LDNHYLMKMKMYSGCNVLEN 62
Oy      60 LELTYLPNTASLSFLQDIOEVGYVLIANQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
Db      63 LEITYGENDSLSFQSIQEVGYVLIAMNEVSTIPVNLRLIRGONLYEGNFTLLVMEN 122
Oy      120 --GDPLNPNFTVSFMLRVPKVSASHLEQLR--SLTEILKGVLIQRNPQLCYODTILW 175
Db      123 YQKNPS-----PDVYGVGLKQLQLSNLTILSGGVKVSNNPLLCNVETIWM 169
Oy      176 KDIFPHNQLALTLIDTRSRACRCPSCMKGSRWCGSSSECCGLTITVCAGGC--ARCK 234
Db      170 WDIDVKISNPTMNLPLPAHFERQCKQCHGCVNGSCWAPGHCQKFTKLLCAEQGNRRCR 229
Oy      235 GPLPTDCCHQCAAGCTGPKHSDCLACLFNHSIGICELACALVYNTIDTFESMNPGR 294
Db      230 GPKPIDCNEHCAGGCTGPRATDCLACRDFDDGCKXQTCPPKRIYDILVSHOVNPNIX 289
Oy      295 YTFGASCTTACPVNTYLTSDVSGCTLVCPFLHNOETVAEDGTORCEKSRPCARVCYGLME 354
Db      290 YTFGAACVKECPSSNVVYTE--GACVRSASAGMLEVD--ENGKRSCKRCDCVCPVCGIGIG 347
Oy      355 HLREVRATVSANIOEPAGCKKIFGSLAFLPSPFGCDPASNTAPLPDQLOVFTLEETIG 414
Db      348 SLSTTIANSTNINSFSNCTKINDITILNRSFSDPDHYKIGTWDPEHMLNLTVEKETIG 407
Oy      415 YLYISAMPDSLPLDLSVFQNLQVIRGRIILHNGAVS--LTQIGIGISWLGIRSLRELGSLAL 473

```

```

Db      408 YLIVIMWPNMTSLSVFQMLTEIRGRTEFSRGFSFVVVQVRHLQWLGSLSKESVAGNYI 467
Qy      474 IHHNTHLCVHTVHPWQLFRNPHQALHTANRPEDECVBGLACHOLCARGHWGCGPTQ 533
Db      468 LKNTLQLRANTIMRRLFRSEDOISLEYART-----ENQTCNNECSDDGCGWGPPTM 520
Qy      534 CYNCSQFLGEOECVECECVLQGLPREYVNAHCLPCGHECOPONGVTCGPEADQCVAC 593
Db      521 CVSCILHVDKRGKRVASCNLLQGEPRBAQVDRGVQCHQDELTCTDTCGPRAPACSKS 560
Qy      594 AHYKDPFCVARCSPGVKBDLSYMPIKRPDEEGACQPCPCINCTHSCVDLDDKCSRAEQR 653
Db      581 AHFQDQPCICPRCPHGILGSDGDTL-IMKYADKMGQCCQCHQNTQCCSGPGLSGCRGD-1 638
Qy      654 ASPLTISVAVVLCILLVVLGVVFGILIKRQOKIKYTMRLLOTELVETLTPSGAMP 713
Db      639 VSHSSILAVGLVSGLLITVVALIIVLLRRRIK-KRRTIRCLLOEELVEPLTPSGQAP 697
Qy      714 NOAMRILKETELRKVLGSGAFGVYKGIWIPDENVKIPVAIKVLRENTSPPKANKET 773
Db      698 NQAFRLILKETEKORVLSGAFGVYKGLMNPDEGIRIPVAIKVLRATSPKNOEV 757
Qy      774 LDEAVYVAGVSPYVRLIGICTSTVQVLTQIMPGCLLDHVENRGKIGSODLLNWCN 833
Db      758 LDEAVYVAGVSPYVRLIGICTSTVQVLTQIMPGCLLDHVENRGKIGSODLLNWCN 817
Qy      834 QIAKGSYLEDVYLRDLAARVULKSPHVKITDQGLARLDIETENHAGGVYPIK 893
Db      818 QIAKGNVYLEERLVHRDLAARVULKSPHVKITDQGLARLDIETENHAGGVYPIK 877
Qy      894 MMALLESILRRFRTHQSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPORPIC 953
Db      878 MMALLESILQWYTHQSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPORPIC 937
Qy      954 TIDVYIMVYKCMWIDSECRPRFELVSEFSRMAKDRQRFVYVIGNEGLGAPSLDSTFYRS 1013
Db      938 TIEVVYIMVYKCMWIDSECRPRFELVSEFSRMAKDRQRFVYVIGNEGLGAPSLDSTFYRS 994
Qy      1014 LLEDDMDLVDAEELVLPQGGFFCPDRPARGAGGVVHNRSSSTSGGDDLGLGLEPSE 1073
Db      995 LLSDDP-DVVDDEVELLPYKRI-----NNGQS----- 1020
Qy      1074 EEARSPPLARSEGAGSDVFDGDLGMGAKGLQSLPTHPSPRLORYSEDPV-PLPSETDG 1132
Db      1021 -----EPCIPPTGH-----PRENSITLRNISDPTQMLKXDLG 1055
Qy      1133 YVAPRLTCSPOPEYVNOQVRRP-----PSPRE-----GLP-AARPGATLEPAKTL 1179
Db      1056 H-----EYVNOQVSETSRSLSDIYNPNYEDLTDGMPVLSLQEAETVPSRPEYL 1105
Qy      1180 SPQKNGVVKDVFAGGAVENREYLTPQCGAARPHRPAASPAFDNIYVNDQDPRPERGAR 1239
Db      1106 NTNQNSL-PLVSSGMDPDY--QAG-----YQAAF-----LPQTAL 1140
Qy      1240 PSTFKCTPTAENPEYIG 1256
Db      1141 TGNQMFLLPAENLEYIG 1157

```

```

RESULT 9
ERR3_HUMAN
ID ERR3_HUMAN STANDARD; PRT; 1342 AA.
AC P2160;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Receptor protein-tyrosine kinase exBb-3 precursor (BC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Porgescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]_
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Ploman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]_
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS. A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE p85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR GeneW; HGNC:3431; ERBB3.
DR MIM; 190151; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PR00069; pkinase; 1.
DR Pfam; PR00757; furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

```

KM	Alternative splicing.	
FT	SIGNAL	1 19
FT	CHAIN	20 1342
FT	DOMAIN	20 643
FT	TRANSMEM	644 664
FT	DOMAIN	665 1342
FT	DOMAIN	709 966
FT	NP BIND	715 723
FT	BINDING	742 742
FT	ACT SITE	834 834
FT	DISULFID	186 194
FT	DISULFID	190 202
FT	DISULFID	210 218
FT	DISULFID	226 226
FT	DISULFID	227 235
FT	DISULFID	231 243
FT	DISULFID	246 255
FT	DISULFID	259 266
FT	DISULFID	290 301
FT	DISULFID	305 320
FT	DISULFID	323 327
FT	DISULFID	500 509
FT	DISULFID	504 517
FT	DISULFID	520 529
FT	DISULFID	533 549
FT	DISULFID	552 565
FT	DISULFID	556 573
FT	DISULFID	576 585
FT	DISULFID	589 610
FT	DISULFID	613 621
FT	DISULFID	617 629
FT	CARBOHYD	126 126
FT	CARBOHYD	250 250
FT	CARBOHYD	353 353
FT	CARBOHYD	408 408
FT	CARBOHYD	414 414
FT	CARBOHYD	437 437
FT	CARBOHYD	469 469
FT	CARBOHYD	522 522
FT	CARBOHYD	566 566
FT	CARBOHYD	616 616
FT	VARSPLIC	141 183
FT	VARSPLIC	184 1342
FT	CONFLICT	560 560
FT	CONFLICT	1064 1064
SO	SEQUENCE	1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
Query Match	35.5%; Score 2427.5; DB 1; Length 1342;	
Best Local Similarity	40.5%; Pred. No. 8.1e-123;	
Matches 534; Conservative 192; Mismatches 453; Indels 141; Gaps 34;		
QY	10 GLLALLPFGAA--STVCTGTDMLRLPASEPHTLMDLRLHYOGGQVQVQNLLEYLPT	67
DB	11 GLTSLRAGSEVGNQAVCPGTLNGLSYGAENQYQTLKYKLERCEVVMGMLLEYLVTSH	70
QY	68 NASISFLDIOEVQVYLIAHQVROVPLQRLRIYRGTLQFEEDNVALAVLDMDLNNN	127
DB	71 NADLSFLDIREVTVYLVANNEFSTLPLNLRVARGVGVGKRAIFV-----MLNTN	124
QY	128 NFTVSFWLRVPRKVSASH-LEQLR--SLTEILKGVVLQIRNPQLCYQDTLLMKDIFHKNO	184
DB	125 -----TNSSHLRQLRLTQVTEILSGVYIEKNDLCHMDTIDWRDIDVRDR-	171
QY	185 LALTLIDTNRSLACHPGSPMKSGKMGESSDQSLRTTVAGGC-ARCKPRLPTDCH	243
DB	172 --AEIVLVNDNGSCPCPCHEVCKG-RCWGGSSEDCTLTCTICAPCNGHCFGPNQOCH	228
QY	244 ECGAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDPFESNPNEGRTFGASCYT	303
DB	229 DECAAGCGGPDQTDCAFCRHFNDSGACVPRCPQLVYNKLTIFOLEBNPTTKYQYGVCVA	288

QY	304 ACPYNYLSTDVGSCTLVCPRLHNOETABDGTORCEKSKRCARVCYGLGMEHLREYAVT	363
DB	289 SCPHNHV--VDQTSVCRAACPRDKMEVD-KNGLMKCEPCGGLCPRACGTGSG--SRGTVD	344
QY	364 SANIOEFACCKKIFGSLATLPESFGDDPASMNTAPLOPOLQVLETLEYITGVLYISAMP	423
DB	345 SSNIGFVNCITLGLDELFTLGLNDGPMHKIPALDPEKLNFRVREITGVYINISWPP	404
QY	424 SLPLDSVPONLQVINGRILHNGAYS-LTLQGLISMLGRLSRELQSLALIHNNHLCF	482
DB	405 HNNHNSVFENLTTTIGRSLYNGFSLINKNLNVTSLSGRSLKEISAGRIYISANQLCY	464
QY	483 VHTVPMDOLEFRPHOALLATA-NRPEDECVGEGLACHOLCARHNCMPPTQVNCNSQFL	541
DB	465 HHSIMMTKYLRRPTEBERLDIKNNRPRRDCVAAGKVCDPICSSGCGCPGPGCLSTRNYS	524
QY	542 RQECVEBCRVLOGLPREYVNAHCLPCHEPCOPONGSVTCGPREADQCAAHYKDPF	601
DB	525 RGVCVTHCNFNLNGEREFRAHEACFSCHEPCOPMEGTATCNGSGSDTCAQCAMFRDGRH	584
QY	602 CVARCPGSKPRLSYMPIKPFDEBGAOCPPCINCHSCVDLDDKCCRAEGR-SP	657
DB	585 CVSSCPHGLG--AKGPIKYPDVQNECRPCHECTQCGKGPLODCLQTLVLIGKTHL	642
QY	658 TSIVSAVVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOSTLEVEPLTPSGAMPNOA	716
DB	643 TVALYVING--LVVIFMMLGTFVWGRGRIONKAMRKYLERGSEITEPLDPS-ERANKV	699
QY	717 QMRILKETELRNVKVLGSAFGSTVYKGIWIPGSENVKIPVAIKVLENTSPKANKEILDE	776
DB	700 LARIFEKTELRLKVLGSGVFPCTHKGVMIPEGESIKIPVCIKIVIEDKSGRQFOAVTDH	759
QY	777 AVYMAQVSPYVSLIGTCLTSTVOLYQOLMYGCLLDHYRENRGLSGODLLMKMOIA	836
DB	760 MLATSLDHAIVRLGLCPGSSLOLVQYLPGLSLDVRHGRGLAQPLLNMVQOIA	819
QY	837 KMSYLEDVRLVHRDLAARNVLYKSPNHKIIDTDFGLARLIDIDETEYHADGKVPYIKMA	896
DB	820 KMYVLEBGMVHRDLAARNVLYKSPQVQVADLPDDKOLLXSEAKTPYIKMA	879
QY	897 LESILRRFTHOSDVWSGVTVWELMTFGAKPYDGI PAAREIDPLEKGRLEPQPICTID	956
DB	880 LESIHFGKXTTHOSDVWSGVTVWELMTFGAEPYAGIRLAEVPLDLEKGRLAQPICTID	939
QY	957 VYNIWVKCMIMISECRPRRELVSFESRAARPORFVYIQNEDLGRA---SPIDSTFYYS	1013
DB	940 VYMWVWKCMIMIDENIRPTFEKLANEFTRMARDPPRYLVIKRES-GPGIAPGEPHGLTYNK	998
QY	1014 LLEDDMDGLVDAEYLVPOQGFCCPDRPARGAGVNHHRSSSTRSGGDLTLGLEP-S	1072
DB	999 KLEVELEFELDLDLDLEAEED-----NLATTTLSALSPLVGTIN	1039
QY	1073 EEEARSPPLAPSEGASDVFDGLGMGAQGLQSLPTHD-PSPLORYSDDPTVPLP----	1127
DB	1040 RPRGQSILSPSSGY-MPMNQNLGESSQESAVSSGSECRPRVSLH-----PMRGLCL	1092
QY	1128 --SETDGYA-----PLTSGQPE-----YNOQPDVPRPPSPBRSP----	1162
DB	1093 ASESSEGHVTSGEALOEKVMKCRSRSSRSRPRGDSAYHQSHSLTLPVPLSPGLE	1152
QY	1163 -----LPAARPAATLERAKTLS-P-GKNQV-----KQVFAFGAVENPEYLTPOG	1207
DB	1153 EEDVAVGYWPDHLNAGTSSREGTSSVGLSSVLTGEEDD-----EEYVNNRR	1204
QY	1208 GAAPRPPRPAFSPAFDNLVYWD-----QDPERGAPSTFGKPTLAENPEYL	1255
DB	1205 RHSP-RHPPRPSLELEGEYMDVDSLSASLGTSQSCPLHVPRLMPTAGTTPDDDEYEM	1263
RESULT 10		
ERR3_RAT	STANDARD;	PRT; 1339 AA.
ID	ERR3_RAT	
AC	Q62759; Q62955;	

DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)  
DE (c-erbB3).  
GN ERBB3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OK NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=96096535; PubMed=8522190;  
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
RT "Cloning of the rat ErbB3 cDNA and characterization of the  
RT recombinant protein.";  
RL Gene 165:279-284(1995).  
RN [2]  
RP REVISIONS TO 85; 513 AND 565.  
RA Hellyer N.J., Koland J.G.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 922-1097 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
RX MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neueregulin and their putative receptors, ErbB2 and  
RT ErbB3, is induced during Wallerian degeneration.";  
RL Neurosci. 17:1642-1659(1997).  
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.  
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC - SUBCELLULAR LOCATION: Type I membrane protein.  
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
CC PHOSPHATIDYLINOSITOL 3-KINASE.  
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U29339; AAC28498.2; -;  
DR EMBL; U52530; AAC53050.1; -;  
DR HSSP; P11362; IFCG.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00219; TYKIC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR TrEMBL; P50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
FT CHAIN 1 19  
FT SIGNAL 20 1339  
FT DOMAIN 20 643  
FT EXTRACELLULAR (POTENTIAL).  
FT

FT TRANSMEM 644 662  
FT DOMAIN 663 1339  
FT DOMAIN 183 259  
FT DOMAIN 707 964  
FT NP\_BIND 713 721  
FT BINDING 740 740  
FT ACT\_SITE 832 832  
FT DISULFID 186 194  
FT DISULFID 190 202  
FT DISULFID 210 218  
FT DISULFID 214 226  
FT DISULFID 227 235  
FT DISULFID 231 243  
FT DISULFID 246 255  
FT DISULFID 259 266  
FT DISULFID 290 301  
FT DISULFID 305 320  
FT DISULFID 323 327  
FT DISULFID 500 509  
FT DISULFID 504 517  
FT DISULFID 520 529  
FT DISULFID 533 549  
FT DISULFID 556 573  
FT DISULFID 576 585  
FT DISULFID 589 610  
FT DISULFID 613 621  
FT DISULFID 617 629  
FT CARBOHYD 126 126  
FT CARBOHYD 250 250  
FT CARBOHYD 353 353  
FT CARBOHYD 408 408  
FT CARBOHYD 414 414  
FT CARBOHYD 437 437  
FT CARBOHYD 469 469  
FT CARBOHYD 522 522  
FT CARBOHYD 566 566  
FT CARBOHYD 616 616  
FT CONFLICT 1028 1028  
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CMC64;  
Query Match 34.4%; Score 2357.5; DB 1; Length 1339;  
Best Local Similarity 40.7%; Pred. No. 4,6e-119;  
Matches 526; Conservative 173; Mismatches 426; Indels 167; Gaps 36;  
QY 3 LALCEWGLLLLPFGA---STVCTGDMKLRIPASETHLDMRLHYCCOVQGN 59  
DB 7 LQVLC---FLSLARGSEMSQAVCPGLNGLSTGADNOYQTLTKYKCEVVMGN 62  
QY 60 LELTYLPTNASLFLDIOEVQGYVLIAHQVQVPLQRLIVRGTLFEDNYALAVLDN 119  
DB 63 LELVLTGHNADSLFLQWIREVTGYLVANNEFSVPLPRLIVRGTVGDKFAITV--- 119  
QY 120 GDPPLNNFNFTVSEFWLVRKVSASH-LEQLR--SLTEILKGVLIGRNPOLCYODTTLWK 176  
DB 120 ---MLNLYN-----TNSSHALRQLKFTLTELISGQVYIEKNDLCHMDITDMR 164  
QY 177 DLFHKNQLATLIDTNRBRACHPCPMKSGKSCWSESSEDCSLRTVCAGGC-ARCKG 225  
DB 165 DIVRRV---GAEIVVKNNGANCPRCHEVKG-RCWGPBGDDCQILTKTICAPQNGRCFG 220  
QY 236 PLPTDCHEQCAAGCTGPRGSDCLALFHNSGICLGHAPALVTYNTDFESPNREGRY 295  
DB 221 PNPNGCHDECAAGCGGPGDTCFCACRRRNDGACVPRPREPLVYVKLTFGLEPNHTKY 280  
QY 296 TTGASCVTACAPNYVSTDVSGCTVPLRNLQEVTAEDGTQRCSEKSPCARVCYGLAMEH 355  
DB 281 QYGVGVVASCPRHFV-VDTQFCVRACPRPKMEVD-KHGKMKESPCCGLCPKACEGSG- 337  
QY 356 LREVRVATSAIOEFAGCKKIGSLAFLPESPDGDPASTATPLQPEQLQVFTLEITGY 415  
DB 338 -SRVQTVDSNIDGFNCTKILIGNDPLFTGLAVDPMHKI PALDEPKLVNFRVIRITGY 396  
QY 416 LVYSAMPDSLPLLSVFPQNLQVIRGLIHNGAYS-LTLQIGISWLGRLSRLEIGSGIALI 474

```

Db 397 LNIQSPHMHNFVSFSLTTIGSLYNRGSLIMKNLNTSLGFSLEISGRVYI 456
Qy 475 HHNTHLCFVHTVPMDFRNPHQALHTA-NRPEDEVCEGLACIQLCARGHCMGPGPTQ 533
Db 457 SANQOLCYHHSLNMTLRRLRGPSEERLDIKYDRPLCECLAEGRVCDPLSSGGCKPBGQ 516
Qy 534 CVNCSQFLRGCEVCEECVLOGLPREVYNARHLCPCHEPCOPONGSVTCFGEADQVAC 593
Db 517 CLSCRNVSRGVCTVHCNCFLOQEPREFVHAQCFSCHEPCECLMEGSTICNGSGDACARC 576
Qy 594 AHYDPRFCVARGSGVPRDLISYMPIKPEDEEGACQCPINCHTSC--VLDLDGKCAE 651
Db 577 AHFRDGRHCNVCSPHIG--AKGTYKYPDAQNECRCHENCTGCGNPELDGLQAE 634
Qy 652 QRASPLTSIVAVGILVVLGVVFGILIKRROKIR-KYTMRLLOETELVEPLTPSG 710
Db 635 VLKSPHVIIVATVQ--LAVILMLGGSFLVWRGRIONKAMRYLERGESIEPLDS- 691
Qy 711 AMPQOAMRILKETELRKVVLGSGAFVTYKGIWIPGENVKIPVAIKVIRENTSPKAN 770
Db 692 EKAKVKVARIKETELRKVLGSGVFGTVHKGIWIPGESIKIPVCIKVIEDKSGRQSF 751
Qy 771 KEILDEAVMAGVSGSPVYSLGICLTSTVOLVQMLPYGCLDHYRNRGRGLGQDLIN 830
Db 752 QAVIDHMLAVGSLDHAHIVRLGLCPGSSLDLVQYLPGLSDHVKOHRRTGLPOLLN 811
Qy 831 WCMQIAGKMSYLEDVRLVHRDLAARNVLYKSPNHVKTDFGLARLLDIDETEHYADGKV 890
Db 812 WGVQIAGKMYLLEHSHVHRDLARNVLYKSPQVADDFVADLPPDDKQLHSEAKT 871
Qy 891 PIKMALESILRRPFTHOSDVSYGVTVWEIMTFGAKPYDGI PAEIPDLLEKGERLPOP 950
Db 872 PIKMALESIFGKYTHOSDVSYGVTVWEIMTFGAEYAGIRLAIEIPDLLEKGERLPOP 931
Qy 951 PICTIDVYMIWVKCMIDSECRPRELVSEFSRARPQRFVYIQNDLDPASGLDSTF 1010
Db 932 QICTIDVYMWVKCMIDENIRPTFKELANEFTRAARPPRLVYIKRAS -PGRTP--PAA 988
Qy 1011 YRSILDEDDMDGLVDAAEYLVPOQGFPCPDPAFCAGVWHHRHRSSTRSGGDLTLGLE 1070
Db 989 EPSVLTKTEL-----QEALELEPL-----DIDDLLE 1014
Qy 1071 PSEF-----EAPRPLAPSEG-----AGSDVFPDGLMGAAK 1102
Db 1015 AEEEGLATSLGALSLLPTGTLTRPRGSQLSPSSGYPMNQSIGEACLDSAVLGSREQ 1074
Qy 1103 GLOSLPHDPSPLRYSRDPVPLPSERDGV-----ATL-----TC-----SPOPE-- 1144
Db 1075 FSRPISLH-PIPRGR-----PASESEGHVTSSEALQKVSVCRSRSRSPRPRGD 1126
Qy 1145 --YVNDPVRPQPSPREGP-----LPAARPAATLERAKTLSF-GKQGVV--- 1167
Db 1127 SAHNSQHSLLTPTPLSPLEEDDNGYMPDTHLGRASSREGTILSSVGLSLVLTGE 1186
Qy 1188 ---KDVFAFGAVENPEYLPFGGAAPQHP 1216
Db 1187 EEDED-----EEYEYMKRKRGRSP-PRPP 1209

```

## RESULT 11

```

ID EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (EGfr)
DE (Gurken receptor) (Tropo protein) (Drosophila relative of ERBB).
OS EGFR OR TOP OR C-ERBB OR DER OR CG10079.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

```

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schupbach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlusser J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Buttis K.C., Buesam D.A., Butler H., Cadieu E., Cemer A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Hariri N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalalet M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matvei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

```

RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 RN [7]  
 RP SEQUENCE OF 959-1078 FROM N.A.  
 RC STRAIN=Daekwanryeong;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.,  
 RT "A *Drosophila* genomic sequence with homology to human epidermal  
 RT growth factor receptor,"  
 RL Nature 314:178-180(1985).  
 RN [8]  
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.Z.,  
 RT "Interallelic complementation among DER/Flb alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases,"  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RP REVIEW.  
 RA MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.,  
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*  
 RT EGF receptor,"  
 RL Cell 89:13-16(1997).  
 CC -I- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES. AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTECOSA  
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -I- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III: ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -I- TISSUE SPECIFICITY: UNBOUTHOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF052754; AAC08536.1; -;  
 CC EMBL; AF052753; AAC08536.1; JOINED.  
 CC EMBL; AF052754; AAC08536.1; -;  
 CC EMBL; AF052752; AAC08535.1; JOINED.  
 CC EMBL; K03054; AAAS1462.1; -;  
 CC EMBL; K03417; AAAS1460.1; -;  
 CC EMBL; K03416; AAAS0965.1; -;  
 CC EMBL; K03418; AAAS1461.1; -;  
 CC EMBL; AF109077; AAD26134.1; -;  
 CC EMBL; AF109078; AAD26132.1; -;

DR EMBL; AF109082; AAD26132.1; JOINED.  
 DR EMBL; AF109078; AAD26133.1; -;  
 DR EMBL; AF109094; AAD26133.1; JOINED.  
 DR EMBL; AF109079; AAD26130.1; -;  
 DR EMBL; AF109081; AAD26130.1; JOINED.  
 DR EMBL; AF109079; AAD26131.1; -;  
 DR EMBL; AF109083; AAD26131.1; JOINED.  
 DR EMBL; AF109080; AAD26135.1; -;  
 DR EMBL; AE003454; AAP46732.1; -;  
 DR EMBL; X02293; CAA26157.1; -;  
 DR EMBL; X78920; CAA55523.1; -;  
 DR EMBL; X78918; CAA55521.1; -;  
 DR EMBL; X78919; CAA55522.1; -;  
 DR PIR; A00640; GOFPE.  
 DR HSSP; P11362; TECK.  
 DR FLYbase; FBgn003731; Egfr.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00069; kinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recp\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SMO0261; FU; 7.  
 DR SMART; SMO0219; TYKIC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;  
 KW Developmental protein.  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 869 1426 POTENTIAL.  
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 938 1198 PROTEIN KINASE.  
 FT NP\_BIND 944 952 ATP (BY SIMILARITY).  
 FT BINDING 971 971 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1063 1063 BY SIMILARITY.  
 FT MOD\_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 Query Match 28.9%; Score 1977; DB 1; Length 1426;  
 Best Local Similarity 32.9%; Pred. No. 1.2e-98;  
 Matches 470; Conservative 185; Mismatches 431; Indels 344; Gaps 42;  
 Oy 24 QVCTGTDMKRLPASPETHLDMLRHLVYGGQVQGNLELYLPT-NASLSFLDIOEVOG 82  
 Db 100 KICIGTGRSLVSPSKHEHYRNLDRTYVNDGKLTWLPENLDLSFLDINIREVTG 159  
 Oy 83 YVLAHNVQVRPLQRLRIVRGTOF-----EDNYALAVALNDGPLANNFNFTVSFWLRV 137  
 Db 160 YILISHVDVKKVPKQIIRGLTFLSVSEKVAL-----FVYTSKMYTLEI 208  
 Oy 138 PKVASHLEQLRSLEILKGVLLQRPQLCYODTILMKDIFKRNQALALLIDNRRA 197  
 Db 209 P-----DLRDVNLQGVGFHNNYNLCHMRITQMSSEIVNSGTDAYVYVDFAPBRE 257  
 Oy 198 CHPCSPMKGRGCGSESEDDCSLTRVCAGACA--RCKGPLDPDCHGEOCAAGTGPKH 255  
 Db 258 CPKHESCTHG-CHGEGPKNCKQSKLCSQACAGRGYGRKPRECHLPACAGGTGPTQ 316  
 Oy 256 SDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACENYVLTSDVG 315  
 Db 317 KDCIACKNFPEDEAVSKCECPMRKYNPTTYVLETNPEKGAVYAGATCVKECP-GHLLRDNG 375  
 Oy 316 SCTLVCPPLHNOEVALEDTQCEKSCPKCAVCGLGHEHLRBYRAVTSANIOEAGCK 375  
 Db 376 ACVRSCPDQKMDKGE-----CVPCKGCPPTCGGVYVLA-----AGNIDFRNCTV 422  
 Oy 376 IFGLAFLPESFDG--DPASNTA-----PLQRPQLQVFETLEITGYLYISAPDSDLPD 427

```

423  IDGNRIIDQTSFGQDVYANTWGPRIYPLDPEREVSTYKETGTGYNIEGTHPQREN 482
428  LSVFONLQVIRGIRILANGAY-SLTLOGLIGISWLGRLSELGSGALIHNTLCEVHTV 486
483  LSYFNLELTHROQLMESWFAALAIYKSLSYLEHRLKQISSGAVIOHNDLQVSNII 542
487  PWDOLFRNPQALHTANRPEDECVEGLACHQLCARGCWCPGPTQCVNCSQPLRGQEC 546
543  RMPALOKEPEOKVWVNENIRADLCENKGTICSDQCNEDGCMGAGTDQCLTCNENFNCTC 602
547  VEEGCVLQGLPREYVNAHCLPCHPECCQONSGSVCFGEAOQCAVACAHYKPPFCVAC 606
603  IADCGYISNAYK--FDNRTCKICHEPCR-----TCNGAGADHCOECVAVRDGHCVCSEC 654
607  P-----SGVK-----PDL 614
655  PKNKXNDRGVCRECHATCGCPKPDITIGACTTCNLAIIINDATVRCCLKDKDCPD- 713
615  SYMPLWKF--PDEBACOP-----CPI-----NCTH----- 638
714  GY--FMEYVHPQEQSLKPLAGRAVCRKCHPLCELCNTVGYHEQVCSKCTHYKREQCET 771
639  -----SC-----VDLDDK----- 647
772  ECPADHYDEBQRCFQHRHPECNCTGCGEADDCSCRNFKLPDANETGYVNSTMFCNTS 831
648  -CPAEQR-----ASPLTS-----IYSAVAGILVVLGVFGI 679
832  KCPLEMRHVNQYTAIGPYCAASPSSKITANDVMFIITGAVLVETIILCV--T 889
680  LKRRQOKIRKYT--MRRLQETELVEPLTSGAMPNOQMILKETELRKVKVLGSGAF 737
890  YICROKQAKKKEIVMTALSGCEDESEPLRPSNIGANCKIRIVDAELRKGVLGWGA 949
738  GTVYKGIPIPGENKIVAIKVLAEHNSPKANKKILDEAVYMAVGSPYSRLLGICLT 797
950  GRVYKGVAVPSEENKIVPAIKELKLSGASESEFELEAYMASEEHNVLKLLAYVCS 1009
798  STVOLVTOIMPYGLLDHVRNENRGLSGODLLMCMQIAKMSYLEDVRLVHRDLAANV 857
1010  SOMMLITQMLPGCLLDVYRNNRDKISGKALLNMTQIAKMSYLEEKRLVHRDLAANV 1069
858  LKSPFNHYKITDFGLARLLIDETTYHADGKVPKIMMALESILRRRPTHOSDVSYGVT 917
1070  LVQTPSLVKITDFGLAKLLSDPSNKEYKAAAGKMPKIMLALBECIRNRVFTSKSDVAVGYT 1129
918  VMEIMTFGAKPYDGIAREIDPLEKGERLPQPCITIDVYIMKCMMDISECPRRRE 977
1130  IMELTFQORPHENIPAKDIPDLLEVGLKEQPEICSDICTLLSCWHLDAAMPRTFKQ 1189
978  LVSESRMARDDPQRFVVIQNEDLG--PASPLDSTFYRSLLBEDD--DMGDLVDABEYLV 1032
1190  LTTYPAFAFARDPGRLLAIPGDKFTRLPA-----YTQDEQDLIRKLAPTTDGEALAK 1242
1033  QCGFFCPDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPSEBEAD-----RSLPAPSEG 1086
1243  PDDYLQFPAAGPS-----HRTDCT-----DEMFKLNRKYCDPFSKNSS 1281
1087  AGSDVFDG---DLGMAKAGLSLPTHPSPLQKRSDEPTYPLBSETGYVAPLTCSPQ 1143
1282  TGDDERDSSAREVGGNLR-----LDLPDEDDYDIMP--TCQGP 1319
1144  EYVNPQDVRPOPSPREGPLPAARPAAGATLBRAKTISGKGVVDVPAFGAAYNPXYL 1203
1320  NNNNNNN-----NPNQNNMAAVGAAGYM-----DLGVPVSDNPEYL 1358
1204  -----TPQGAAPQPH-----PPAFSP-AFDNLYYWD 1230
1359  LNAQTLGVGSPPIPTOTIGIPVMGPGTMEYKVPMPGSEPTSSDHEYYND 1408

```

RESULT 12

```

ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OC NCBI_TaxId=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8522822; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor".
RL Cell 41:719-726(1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INDUCED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M10066; AAA48763.1; ALT_INIT.
DR PIR: A00643; TVCHLV.
DR PIR: B00643; TVFVLV.
DR HSSP: P1362; IFGK.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; TyKc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KM Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146
FT BINDING 165 165
FT ACT SITE 257 257
FT SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 25.6%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 7,6e-87;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
OY 593 CAHYKDPFCYARCPGKPLSLWPKFPEBEGACQPCPINTGHSCTVDLDDKQCPAQ 652
DB 3 CAHFIDGPHCVKACPAVAGENDTL-VWKYADANAACVQLCHPNCRTGCKGPGLECCP--- 58
OY 653 RASPLTGISVAVV-GILLVVLGVVFGILIRROOKIRKYMRRLQETELVEPLTSPGA 711
DB 59 NGSKTPSIAGVAVGGLCLLVNAGIGLYLRR-HYVAKRTLRLQERLVEPLTSPGE 117
OY 712 MPNOQMILKETELRKVKVLGSGAFVYKGIPIPGENKIVPAIKVLRENTSPKANK 771

```



```

DB      118 AENOAHRLIKETEFKVKVVLGSGAGFTYKGMIPGEKVKIPVAIKELREATSPKANK 177
QY      772 ELIDEAYVWAGSPVYSRLGICLTSTYQVLTQMLPYGCLDHWENGRGLGSDLLMW 831
DB      178 ELIDEAYVWASVDNPNVCHLCITSTYQVLTQMLPYGCLDHWENGRGLGSDLLMW 237
QY      832 CMQIAKAGMYLEDVRLVHRDLAARNVLYKSPNHVKITDPGLARLDIDETEHYADGKYP 891
DB      238 CQIAKAGMYLEERLVRHDLAARNVLYVTPGHVKITDPGLAKLGADEKEHYAEGKYP 297
QY      892 IKMALESLIRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951
DB      298 IKMALESLIRHRYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSYLEKGERLP 357
QY      952 ICTIDVYIMVWCMWIDSECRPFRELVSFSPRMADPQRFVYIQ-NEDLGASPLDSTF 1010
DB      358 ICTIDVYIMVWCMWIDADSRPFRELVSFSPRMADPQRFVYIQDERMHLPSPTDSKF 417
QY      1011 YRSLEDDMDGLVDAEYLVLPQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLE 1070
DB      418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY      1071 PSEEPAPSPFL-----APSEGASDVFPDGLGMAKAGLQSLFTHDPSLQRYSEDPTVP 1125
DB      450 -----SRTPLLSLSATSNMSATNCID-----RNGGCHVREDSPQRYSSDPTGN 495
QY      1126 LPSET--DGTVAPLTCSPQPEYVNOPDVRPOPSPREGPLPAARPAAGATLERAKTUSPOK 1183
DB      496 FLEESIDDFL-----PAPEYVNO--LMPKKS-----TAMQO 526
QY      1184 NGVVKDVF-----AFGGAIVENPEYLTTPQGAAPQPPPAFSPAFDNLTY 1228
DB      527 NQIVNNISLTATSKLPMDSRYQNSHSTAVDNPETL-----NTNQSPFLAKTVFESSPY 578
QY      1229 WQO-----DPEP-----RGAPSPFKGTPTEANPEYLGDLVP 1260
DB      579 WIQSGNHQIINDNPYQODFLPNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER
ID_ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00353;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserre G., Claverie J.-M.,
RA Sautie S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.

```

```

CC      -I- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC      -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR      EMBL; K02006; AAA42394.1; ALT_INIT.
DR      EMBL; K01216; AAA42400.1; -.
DR      PIR; A00644; TVYU.
DR      HSSP; P13362; 1FGK.
DR      InterPro; IPR000719; Euk_pkinase.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00069; pkinase; 1.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00219; TyrKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      TRANSFERASE; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132 399
FT      NP_BIND 138 146
FT      BINDING 165 165
FT      ACT_SITE 257 257
FT      CONFLICT 29 29
FT      CONFLICT 140 140
FT      CONFLICT 146 146
SQ      SEQUENCE 604 AA; 67633 MW; 76BECDD06745D009 CRC64;

Query Match      24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 2, 2e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY      593 CAHYKDPFCVACRPSGVKPDLSYPIWKFPPBEGACQPCPINCHSCYVDLDDKGPAPQ 652
DB      3 CAHFIDGPHCVACRPGVIGENDTL-VRRYADANAVCOLCHPCTRGCGPGLGCP--- 58
QY      653 RASPLTISYAVV-GILVWVGVVFGILIKRROOKIRYVTRRLLOETELVEPLTPSGA 711
DB      59 NSKTPSIAGVAVGGLCVVVGIGLVLRRR-HYRKRTLRLLQDELEVEPLTPSGE 117
QY      712 MENOAGMRLIKETELAKVYVLSGAGFTYKGMIPGEKVKIPVAIKELREATSPKANK 771
DB      118 AENOAHRLIKETEFKVKVVLGSGAGFTYKGMIPGEKVKIPVAIKELREATSPKANK 177
QY      772 ELIDEAYVWAGSPVYSRLGICLTSTYQVLTQMLPYGCLDHWENGRGLGSDLLMW 831
DB      178 ELIDEAYVWASVDNPNVCHLCITSTYQVLTQMLPYGCLDHWENGRGLGSDLLMW 237
QY      832 CMQIAKAGMYLEDVRLVHRDLAARNVLYKSPNHVKITDPGLARLDIDETEHYADGKYP 891
DB      238 CQIAKAGMYLEERLVRHDLAARNVLYVTPGHVKITDPGLAKLGADEKEHYAEGKYP 297
QY      892 IKMALESLIRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951
DB      298 IKMALESLIRHRYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSYLEKGERLP 357
QY      952 ICTIDVYIMVWCMWIDSECRPFRELVSFSPRMADPQRFVYIQ-NEDLGASPLDSTF 1010
DB      358 ICTIDVYIMVWCMWIDADSRPFRELVSFSPRMADPQRFVYIQDERMHLPSPTDSKF 417
QY      1011 YRSLEDDMDGLVDAEYLVLPQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLE 1070
DB      418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY      1071 PSEEPAPSPFL-----APSEGASDVFPDGLGMAKAGLQSLFTHDPSLQRYSEDPTVP 1125

```



```

Db      450 -----SRPPLLSLSATSNNSNATCID-----RNGQGHPRVDESFVQRYSSDPNGN 495
Qy      1126 LPSET--DGYAAPLTCSPQPEVYVNOVDYRPPQSPRECEGLPARPAGATLERAKTLSFGK 1183
Db      496 FLEESIDGFL-----PAPEVYVNO--LMPKKESTAM-----524
Qy      1184 NGVAVDVFAP-----GCAVENPEYLTPOGGAAPQHPPEAPSPAPD 1224
Db      525 --VQNGQINFIISLTLAISKLPMDSRQNSHSTAVNDPEYL-----NTNOSPILAKTYFE 574
Qy      1225 NLYYWDQDPPERGAPSPFTFKGTPTAENPEY 1254
Db      575 SSPYWIQSGNHQ-----INLNDPEY 594

RESULT 14
ERBB_AVIEU
AC      P11273;
DT      01-JUL-1989 (Rel. 11, Created)
DT      01-JUL-1989 (Rel. 11, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN      v-ERBB.
OS      Avian erythroblastosis virus (strain ts167).
OC      Viruses; Retrocid viruses; Retroviridae; Avian type C retroviruses.
OX      NCBI_TaxID=1031898;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87064458; PubMed=2878364;
RA      Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT      "A single amino acid substitution in v-erbB confers a thermolabile
RT      phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT      cells.";
RL      Mol. Cell. Biol. 6:1751-1759(1986).

-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
-!- tyrosine phosphate.
-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, M13179; AAA42401.1; -
DR      PIR, A25231; TVPEVB.
DR      HSPR, P11362; IFGK.
DR      InterPro: IPR000719; Euk_pkinase.
DR      InterPro: IPR001245; Tyr_pkinase.
DR      Pfam: PF00069; pkinase; 1.
DR      Prodom: PD000001; Euk_pkinase; 1.
DR      SMART, SM00219; TyrcK; 1.
DR      PROSITE, PS00107; PROTEIN KINASE ATP; 1.
DR      PROSITE, PS00109; PROTEIN KINASE TYR; 1.
DR      TRANSFERASE, Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132 399
FT      NP_BIND 138 146 ATP (BY SIMILARITY) .
FT      BINDING 165 165 ATP (BY SIMILARITY) .
FT      ACT_SITE 257 257 BY SIMILARITY.
FT      VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB) .
SQ      SEQUENCE 540 AA; 60412 MW; 5B53297AA06B65D CRC64;

Query Match: 23.8%; Score 1630; DB 1; Length 540;
Best Local Similarity 53.1%; Fred. NO. 1.6e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
593 CAHYKDPFCVACRCSGVKPDLSWYPIAKPFDEEGACQPCPINCHTSGCVLDLDDGCPAQD 652

```

Dd		3	CAHFLDGHCHYACACBAGVLGENDTL- WIKVADANAANVOQLCHPNCTROCKGGLEGCP	---	58
Qy		653	RASPLTISIVASAV- GILLLVVVGVFGLIKRQOKIRKTYMRRLQETELVEPLTPSGA		711
Dd		59	NGSKTPSIAGAIVGGLLCLVVVGIGIGUYLR-R -HIVRKRTLRRLLQERLEVEPLTPSGE		117
Qy		712	MPNQAKRIILKETELRKRVKLGSAGAFCTYVKGIIRIDGENVKIIVAIKVLRENTSPANK		771
Dd		118	APNQHNLILKETEFKKVKVLFGFAGFYVVGKIWIPEGEKTIIIVALKELREATSPANK		177
Qy		772	EILDAAVMAGVGPVYSRLIGLCTSTVOLVTLOLMYPGCCLDHVRENRGLOGODLNN		831
Dd		178	EILDEAYMASVDNPHRCRLGICLTSTVOLIITOLMFPYGCGLDIIREHKNIISQOYLNN		237
Qy		832	CMOIAKGMSYLEDEVRLVHRDLAARNVLYKSBNHKITDFGLARLLDIDETEHADGGKV		891
Dd		238	CVQJAKGNVYLERRHMVNRDLAARNVLYKPIQHYKITDFGLAQKLGADKEKYHAEGKV		297
Qy		892	IKMMALESILRRRTTHOSDWVSYGVTYWELMTFCASKYYDGI PAEIPLDLEKERLPQP		951
Dd		298	IKMMALESIHRITHTOSDWVSYGVTYWELMTFTCSKPYDGI PASEISSVLEKERLPQP		357
Qy		952	ICTIDVYMIMVKCMNIDSECRPRERELVSFSRRMARPORFVNO--NEDGPSPLDTSF		1010
Dd		358	ICTIDVYMIMVKCMSDDSRKPRRELIASFSSKARPRLVLYIGDERHMLPSPTDSKF		417
Qy		1011	YRSLLDEDMDGLDAEEYLVPOQGFCPPDPAPGAGMVHHRHSSTRSGGDLTLGLE		1070
Dd		418	YRLMEEDMEDVIDADEYLVPHQGF-----NSPST		449
Qy		1071	PSEEARPRSPV----APSBSGSDVFVDGDLGMGAQKLSLPTHDPSPLOYSEDPTVP		1125
Dd		450	-----SRTPLLSSLATSUNNSATNCIDRNCG-----H-----		476
Qy		1126	LPSTTDGVVAPLTCSPQEPYVNQPDVAPOPSPREGPLPARAPAGAT-LERAKTLSPGKN		1184
Dd		477	-PVREDGL-----PAPEVYNQ--LMPKKPSTAMVONQIYVISTAIKSLPIDSKYCN		527
Qy		1185	GIVXDVFAPFGAVENPEYL 1203		
Dd		528	-----SHSTAVDNPEYL 539		
<hr/>					
		RESULT 15			
		EGFR_CHICK			
ID	EGFR_CHICK	STANDARD;	PRT;	703 AA.	
AC	P13387,				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER				
DE	(Fragment).				
GN	EGFR.				
OS	Gallus gallus (chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88261272; PubMed=3260329;				
RA	Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,				
RA	Ulrich A., Vennstrom B., Schlessinger J., Glyou D:				
RT	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,				
RT	expression in mouse cells, and differential binding of EGF and				
RT	transforming growth factor alpha."				
RL	Mol. Cell. Biol. 8:1970-1978(1988).				
CC	-1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,				
CC	AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP130 AND				
CC	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).				
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein				
CC	tyrosine phosphate.				
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.				



; Entered [jdelaval 22-Jul-03 6:11]  
SE04-5-25-12  
MELALCRMGILLALIPGASTOVCTgy.ikanekfigitellYOGCQOVQGNLELYPTNASLFLD  
IOEVGVVLIANNOVROVPLORLRIYRGTOLEFEDNYALAVLDNGDPLNNTTPTYGASPGGLRELORSLT  
ELKGVVLIORNPOLCYODTILMKDIFHKNNOLATLIDNRSRACHPCSPMCKGSRGWSSESDOSIT  
RTVCAGGCARCKGRLPTDCHOCACCTGPKHSDCLACHFNHSGICELHCPALTYNTDTRESNPRE  
GRYTGASCVTACPTNYLSTDVGSCTLVCPHNOEYTAEDGTORCEKSPCARVCYGLMEHLREVRAY  
TSANI OEFAGCKKIFGSLAFLESFODGDPASNTAPLOPEOLOVETLEETGYLYISAMPDSL.PDLVSFO  
NIQVIRGRILHNGAVSLTLOGLGI SWLGRSLRELSGLALIHNNHLCFVHTVPMDOLEFRNHQALHHT  
NARPEDECEGLACIOLCARGHCPGPTQCVNCSQPLRGCEVCECRVLOGLPREYVNAHCLPCHPE  
COPONGSVTCFGEADOCVACAHYKDPFCVACRPSGVKPDLSYMPIMKRPDEGACOPCPINCTHSVD  
LDKCPAROBASPLTSLYSAYVGLLVVVLGVVPGILIKRROOKI RKYTMRLLOETELVEPLTSGAM  
PNOQMRILKETELRKVKVVGSGAFVYKGIWIPDGENVKIPVAIKVLENTSPKANKELDEAVVMAG  
VGSFVYSRLIGICLSTVOLVTOIMPYGCLLDHVRNRRGLSODLLNMCQIAKMSYLEDVRLVHRL  
AARNLVKSPNHWKITDFOGLARLIDIDETEYHADGKVPIMKMALESILRRRFTHOSDWSYGVATWELM  
TFGAPYDGI PAREIDPLEKEGRLEPOPICTIDVYIMVYKMMIDSECPREPRELVSERFSSRA  
VVIQNEEDLGASPLDSTFNRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSG  
GDLTIGLEPSEEEAPSPPLASPEGASDVDPDGLGMAKGLSLPTHDSPLQRYSEDPVPLPSETD  
YVAPLITCSPOPEYVNOVDVPOPPSPRECEPLPARAPAGATLERAKTILSPKNGVVKDVFAPFGAVENPEY  
LTPQGAAPQHPHPAPSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYIGLDVPI

; Entered [jdelaval 22-Jul-03 6:14]  
SE04-5-25-14  
MELALCRMGILLALIPGASTOVCTfnfvcfwLrypkveashlelyOGCQOVQGNLELYPTNAS  
LSRILOIOVOGVVLIANNOVROVPLORLRIYRGTOLEFEDNYALAVLDNGDPLNNTTPTYGASPGGLREL  
OASLTELKGVVLIORNPOLCYODTILMKDIFHKNNOLATLIDNRSRACHPCSPMCKGSRGWSSE  
DCOSLRTVCAGGCARCKGRLPTDCHOCACCTGPKHSDCLACHFNHSGICELHCPALTYNTDTRESNPRE  
SMENPRGRYTFGASCVTACPTNYLSTDVGSCTLVCPHNOEYTAEDGTORCEKSPCARVCYGLMEHL  
REYVAVTSANI OEFAGCKKIFGSLAFLESFODGDPASNTAPLOPEOLOVETLEETGYLYISAMPDSL.P  
DLVSFONLQVIRGRILHNGAVSLTLOGLGI SWLGRSLRELSGLALIHNNHLCFVHTVPMDOLEFRNHQ  
OALHTNARPEDECEGLACIOLCARGHCPGPTQCVNCSQPLRGCEVCECRVLOGLPREYVNAHCLP  
LPCHEPCQONGSVTCFGEADOCVACAHYKDPFCVACRPSGVKPDLSYMPIMKRPDEGACOPCPINC  
THSCVDLDKCGCPAEGRASPLTSLYSAYVGLLVVVLGVVPGILIKRROOKI RKYTMRLLOETELVEPL  
TBSGAMPNOQMRILKETELRKVKVVGSGAFVYKGIWIPDGENVKIPVAIKVLENTSPKANKELDE  
AYVMAGVSPYVSRILIGICLSTVOLVTOIMPYGCLLDHVRNRRGLSODLLNMCQIAKMSYLEDVRLVHRL  
LVHRLAARNVLKSPNHWKITDFOGLARLIDIDETEYHADGKVPIMKMALESILRRRFTHOSDWSYGV  
TYWELMTFGAPYDGI PAREIDPLEKEGRLEPOPICTIDVYIMVYKMMIDSECPREPRELVSERFSSRA  
RDBQRFVVIQNEEDLGASPLDSTFNRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSG  
STRSGGDLTIGLEPSEEEAPSPPLASPEGASDVDPDGLGMAKGLSLPTHDSPLQRYSEDPVPLPSETD  
YVAPLITCSPOPEYVNOVDVPOPPSPRECEPLPARAPAGATLERAKTILSPKNGVVKDVFAPFGAVENPEY  
LTPQGAAPQHPHPAPSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYIGLDVPI

100 100 100  
100 100 100  
100 100 100

: Entered [jdelavel 22-Jul-03 6:21]  
SB04-59-73-12  
MELALICRMGLLALLPFGAASOVCTGTMKRLRPLASPETHLMDLRLHYOGCVOVQGNLELYLPTNAS  
LSFLQDIOEVQVYKANKF1G1E1QRLR1VRGTOLFEEDNYALAVLDNGDPLNNTTPVTGASPGGLREL  
QURSLTEILKGGVLIQIRNPOLCYODITLMKDI FHKNNOLATLIDITNRSRACHPCSPMCKSRCSRC  
DCOSLTRVYAGGACARCKGPLPTDCHEQCAAGCTGPKHSDCLALHFNHSGICELHPALVYNTOTFE  
SMNPEGRYTFGASCVTACVNYLSTVSGCTLYVCPILHNOEVTADGTCORCEKSKCARVCYGLMEHL  
REYRATYSANIOEPAGCKKI FGSIAFLPESFDGDPASNTAPLOPELOVEFTLEITTYLYISA  
DLSVFORNOYIRGRILHNGAYSLTLOGLISWLGRLSRLRELSGLALIHNTHLCFVHTVMDOLFRNPH  
QALLHTANRPEDCEVGEGLACHOLCARGHCHMGPPTOCVNCOSQPLRGQECVEBCRVLOGLPREYVNAHRC  
LPHPEQOPNGSVTCFGEADQVACAHYKDPFCVACRCPGVKPLDSYMPINWKFDEEGACQPCPINC  
THSCVLDLDDKCGCPAEORASPLTSSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPL  
TPSGAMPNOMKRIKETEELRKXVILSGAGTYVKGIMIPDGENVKIPVAIKYLRENTSPKANKETLDE  
AYMAGVGSPTVSRLLGICLTSTVOITOLMPYGCCLDHVRENRGRIGSODLIMWCMQIAKMSYLEDR  
LVHRDLAARVVLKSPNHVKITDPGLARLBDIDETEYHADGKVPPIKMALESILRRRFTTHOSDVMSTYCV  
TVELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICITIDVYIMVWKCMIMDSRCRPRFRELVEFSRMA  
RDQRFVVLQNEBGLSPASPLDSTFYRSLLEDDMDGDLVDAEYVLPQOGFCPPAPAGAGMVHHRSS  
STRSGGDLTLGLEPSEEA PRSPLAPSEAGSDVPDGLGMGAAGLOSILPTHDPSPLORESDPVPL  
PSETDGVVAPLTCSPDEYVNOPOVRPQPSREGPLPAARPAATLERAKTILSGKNGVVKDVFAGGA  
VENPEYLTPOGGAAPQHPPPAPSPAFDNLTYWQDPPERGAPSTFKGTPTAENPEYLGIDVVI

: Entered [jdelavel 22-Jul-03 6:21]  
SB04-59-73-14  
MELALICRMGLLALLPFGAASOVCTGTMKRLRPLASPETHLMDLRLHYOGCVOVQGNLELYLPTNAS  
LSFLQDIOEVQVYKANKF1G1E1QRLR1VRGTOLFEEDNYALAVLDNGDPLNNTTPVTGASPGGLREL  
QURSLTEILKGGVLIQIRNPOLCYODITLMKDI FHKNNOLATLIDITNRSRACHPCSPMCKSRCSRC  
DCOSLTRVYAGGACARCKGPLPTDCHEQCAAGCTGPKHSDCLALHFNHSGICELHPALVYNTOTFE  
SMNPEGRYTFGASCVTACVNYLSTVSGCTLYVCPILHNOEVTADGTCORCEKSKCARVCYGLMEHL  
REYRATYSANIOEPAGCKKI FGSIAFLPESFDGDPASNTAPLOPELOVEFTLEITTYLYISA  
DLSVFORNOYIRGRILHNGAYSLTLOGLISWLGRLSRLRELSGLALIHNTHLCFVHTVMDOLFRNPH  
QALLHTANRPEDCEVGEGLACHOLCARGHCHMGPPTOCVNCOSQPLRGQECVEBCRVLOGLPREYVNAHRC  
LPHPEQOPNGSVTCFGEADQVACAHYKDPFCVACRCPGVKPLDSYMPINWKFDEEGACQPCPINC  
THSCVLDLDDKCGCPAEORASPLTSSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPL  
TPSGAMPNOMKRIKETEELRKXVILSGAGTYVKGIMIPDGENVKIPVAIKYLRENTSPKANKETLDE  
AYMAGVGSPTVSRLLGICLTSTVOITOLMPYGCCLDHVRENRGRIGSODLIMWCMQIAKMSYLEDR  
LVHRDLAARVVLKSPNHVKITDPGLARLBDIDETEYHADGKVPPIKMALESILRRRFTTHOSDVMSTYCV  
TVELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICITIDVYIMVWKCMIMDSRCRPRFRELVEFSRMA  
RDQRFVVLQNEBGLSPASPLDSTFYRSLLEDDMDGDLVDAEYVLPQOGFCPPAPAGAGMVHHRSS  
STRSGGDLTLGLEPSEEA PRSPLAPSEAGSDVPDGLGMGAAGLOSILPTHDPSPLORESDPVPL  
PSETDGVVAPLTCSPDEYVNOPOVRPQPSREGPLPAARPAATLERAKTILSGKNGVVKDVFAGGA  
VENPEYLTPOGGAAPQHPPPAPSPAFDNLTYWQDPPERGAPSTFKGTPTAENPEYLGIDVVI  
VI

; Entered [jdelaval 22-Jul-03 6:37]  
SEQ4-103-117-12  
METALCRMGLIALILPGAASOVCTGDMKLRILPASPETHLDMRLHYOGCOVVOGNLELYPTNNA  
LSFLDIOGVQGVLIJAHNOVQVPIQRLRIYRGTOLEFEDNYALAVLDNGDPLNMFVsfwltvkv  
OARSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOQLATLIDTNRSRACHPCSPMCKGSRCKGSE  
DCQSLRTIVCAGGACARCKPLPTDCCHEOCAGCTGPKHSDCLAHFNHSGICELHCPALVYNTDTE  
SMNPREGRTTFCASCVTACPYNYLASTDVSGCTLVCPAHNOEYTAEDGTORCEKSKPCARVCYGLMEHL  
REVRATVANSIOEPFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVETLEIITGYLXISAMPDSL  
DLVSFONLOVIRGRIIHNQAVSLTLOGIGISMLGRLSRELGSGLALIHNNTHLCFVHTVPMQOLFRNPH  
OALHTARPEDECVGEGELACHOLCARGHCMGPGPTOCVNCOSQFLRGQCEVCECRYLOGLPREYVARRHC  
LPCHCECOPQNGSVTCFGEADOCVACAHYKDPFCVACRPSGVKPDLSYMPIWKPDEEGACQPCPINC  
THSCVDLDKGCPEQRASPLTISVSAVGIILVVLGVVFGILIKRQOKIRKYTMRLQETELVEPL  
TPSGAMPNOAQMRILKETELRKVILGSGAFGVYKGIWIPDEENVKIPVAIKVIRENTSPKXKEILDE  
AYVAGVSPYVSRILGICLTSTVOLVTOIMPYGCLLDHVRNENRGRLGSODLNMCMQIAKMSYLEYV  
LVHRDLAARNVLYKSPNHVKITDFTGLARLLDIDETEHADGKVPIKMMALESIILRRFTHOSDVSYG  
TYWELMTGAKYDGIIPAREIPDLEKGERLPOPPICITIDVYMIWVKCMIIDSECRPRELIVSEFSMA  
RDPORFVUIQNEIDLGPASPLDSTFYRSLLEDMDMDLVDAEYLVPOQGFCCPDPAFGAGVYHHRRSS  
STRSGGDLTTLGEPSEEAAPRSPLAPSEGAGSDVFDGLMGAAKLOSLPTHDPSPLOKRSDEFTVPL  
PSETGVVAPLTCSPQPEYVNOQDVRRPOPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGG  
VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 6:39]  
SEQ4-103-117-14  
METALCRMGLIALILPGAASOVCTGDMKLRILPASPETHLDMRLHYOGCOVVOGNLELYPTNNA  
LSFLDIOGVQGVLIJAHNOVQVPIQRLRIYRGTOLEFEDNYALAVLDNGDPLNMFVsfwltvkv  
saahleQLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOQLATLIDTNRSRACHPCSPMCKGSR  
WGSSSEDCOSLRTIVCAGGACARCKPLPTDCCHEOCAGCTGPKHSDCLAHFNHSGICELHCPALVY  
NTDTPESMNPREGRTTFCASCVTACPYNYLASTDVSGCTLVCPAHNOEYTAEDGTORCEKSKPCARVCY  
LGMEHLREVRATVANSIOEPFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVETLEIITGYLXIS  
WPDLSPLSVFONLOVIRGRIIHNQAVSLTLOGIGISMLGRLSRELGSGLALIHNNTHLCFVHTVPMQ  
LFRNPQALHRTANRPEDECVGEGELACHOLCARGHCMGPGPTOCVNCOSQFLRGQCEVCECRYLOGLPREY  
VARRHCLPCHPECOPQNGSVTCFGEADOCVACAHYKDPFCVACRPSGVKPDLSYMPIWKPDEEGACQ  
PCPINCTHSCVDLDKGCPEQRASPLTISVSAVGIILVVLGVVFGILIKRQOKIRKYTMRLQET  
ELVEPLTPSGAMPNOAQMRILKETELRKVILGSGAFGVYKGIWIPDEENVKIPVAIKVIRENTSPKAN  
KEILDEAYVYVAGVSPYVSRILGICLTSTVOLVTOIMPYGCLLDHVRNENRGRLGSODLNMCMQIAKMS  
YLEYDLVLRHDLAARNVLYKSPNHVKITDFTGLARLLDIDETEHADGKVPIKMMALESIILRRFTHOSD  
VMSYGVTYWELMTGAKYDGIIPAREIPDLEKGERLPOPPICITIDVYMIWVKCMIIDSECRPRELIVS  
EFSRARDQRFVUIQNEIDLGPASPLDSTFYRSLLEDMDMDLVDAEYLVPOQGFCCPDPAFGAGVYH  
HRRSSSTRSGGDLTTLGEPSEEAAPRSPLAPSEGAGSDVFDGLMGAAKLOSLPTHDPSPLOKRSDEFTV  
DPTVPLPSETGVVAPLTCSPQPEYVNOQDVRRPOPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKD  
FAFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPI  
VI

; Entered [jdelaval 22-Jul-03 6:41]  
SEQ4-149-163-12  
MELALCRWGLLALLPRGAASSTOVCTGDMKRLPASPETHLDMRLHLYOGCOVQGNLETTYLPNTAS  
LSFLQDIQEVQGVYLIHNOVROVPLQRLIRVGTQLFEDNYTALAVLDNGDPLNNTTPTVTGASPGGLREL  
QLRSLTEILKGGVLIQNPOLCYODTILMKDQYIkanakIgiLeINRSRACHPCSPWCKSGSRMGESSE  
DCOSLTRTVAGGACARCKPLPTDCHEOCAGCTGPHSDCLACLFHNHSGICELHC PALVYNTDTTFE  
SMNPBGRYTFGASCVTACPVNLTVDGSCCTLVCPILHNOBVTAEADGTORCEKSKPCARVCGYLGMEHL  
REVAVTSANIQEPAGCKKI FGSIAFLPESFDGPASNTAPLOPEQLOVFTLEBTGTYLISAMPSILP  
DLSVFNLOVIRGRILHNGAVSLTLOGIGISWLGRLSRLSGSLALIHNTHLCFVHTVMDQLFNRPH  
QALLHTANRPEDECEVGEGLACHOLCARGHCWPGPTOCVNCOSQFLRGOECVEECRVLOGLPREYVNAHC  
LPCHPEQOPNGSVTCFGEPAQOCVACHYKDPFCVACRCPGVKPDLSYMPIMKFPDEBGAQCPINC  
TSGCVDLDKGCSPAEOBASPLTSIVSAVVGILLVVLGVVFGILLIKRQOKIRKTYMRLLOETELVEPL  
TPSGAMPNQAKOMRILKETELRKXVVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDE  
AYMAGVGSYVSRLLGICLTSTVOLVQMLPYGCLDHYRENGRGLSGODLLNMCQIAGKMYLEDVR  
LVHRDLAARVVLVKS PNHVKITDRLARLIDIDETEHADGKVPIMKMALESILRRRFTHOSDVMSCYV  
TYMELMFGAKPTDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWVKCMIIDSECRPRFRELVEFSHMA  
RDQRFVYIIONEDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGGMVHRHRHS  
STRSGGDLTLGLEPSEEAARSPPLASPEGASDVFDGLGMGAAGLQSLPTHDPSFLQRYSEDPTVPL  
PSETDGVVALTCSPOPEYVNOQPVRRPOPPSPREGPLPAARPAQTLERAKTLSPGKXGVVXKDVAFACGA  
VENPEYLTPOGGAAPQPHPPAFSPAFDNLYYMDODPREGAPSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 6:41]  
SEQ4-149-163-14  
MELALCRWGLLALLPRGAASSTOVCTGDMKRLPASPETHLDMRLHLYOGCOVQGNLETTYLPNTAS  
LSFLQDIQEVQGVYLIHNOVROVPLQRLIRVGTQLFEDNYTALAVLDNGDPLNNTTPTVTGASPGGLREL  
QLRSLTEILKGGVLIQNPOLCYODTILMKDQYIkanakIgiLeINRSRACHPCSPWCKSGSRMGESSE  
DCOSLTRTVAGGACARCKPLPTDCHEOCAGCTGPHSDCLACLFHNHSGICELHC PALVYNTDTTFE  
SMNPBGRYTFGASCVTACPVNLTVDGSCCTLVCPILHNOBVTAEADGTORCEKSKPCARVCGYLGMEHL  
REVAVTSANIQEPAGCKKI FGSIAFLPESFDGPASNTAPLOPEQLOVFTLEBTGTYLISAMPSILP  
DLSVFNLOVIRGRILHNGAVSLTLOGIGISWLGRLSRLSGSLALIHNTHLCFVHTVMDQLFNRPH  
QALLHTANRPEDECEVGEGLACHOLCARGHCWPGPTOCVNCOSQFLRGOECVEECRVLOGLPREYVNAHC  
LPCHPEQOPNGSVTCFGEPAQOCVACHYKDPFCVACRCPGVKPDLSYMPIMKFPDEBGAQCPINC  
TSGCVDLDKGCSPAEOBASPLTSIVSAVVGILLVVLGVVFGILLIKRQOKIRKTYMRLLOETELVEPL  
TPSGAMPNQAKOMRILKETELRKXVVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDE  
AYMAGVGSYVSRLLGICLTSTVOLVQMLPYGCLDHYRENGRGLSGODLLNMCQIAGKMYLEDVR  
LVHRDLAARVVLVKS PNHVKITDRLARLIDIDETEHADGKVPIMKMALESILRRRFTHOSDVMSCYV  
TYMELMFGAKPTDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWVKCMIIDSECRPRFRELVEFSHMA  
RDQRFVYIIONEDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGGMVHRHRHS  
STRSGGDLTLGLEPSEEAARSPPLASPEGASDVFDGLGMGAAGLQSLPTHDPSFLQRYSEDPTVPL  
PSETDGVVALTCSPOPEYVNOQPVRRPOPPSPREGPLPAARPAQTLERAKTLSPGKXGVVXKDVAFACGA  
VENPEYLTPOGGAAPQPHPPAFSPAFDNLYYMDODPREGAPSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 6:47]  
SEQ4-210-224-12  
MEIAALCRMGILLALILPGAASTOVCTGDMKRLPASPETHLDMRLHLYOGCQVNOGNLELTLYPTNAS  
LSFLQDIOEVQGYVLIJAHNOVROVPLORLIRVGTOLFEDNYALAVLDNNDPLNNTTPTVGTASPGGLREL  
OLRLTEILKGGVLIQNNPOLCYODTILMKDI FHKNNQALTLIDTNRSRACHPCSPMCKGSRMGESSE  
DCOSLTRVCAGGACARCKGFLPFIkanrkfigtelkhsdCLACHFNHSGICEHLCPALVTYNTDTE  
SMNPREGRYTFGASCVTACPYNLSTDVGSCTVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHL  
REVRATYSANIQEFACCKIFGSLAPLPSFDDPASNTPLOPEOLQVETLEIEITGYLYISAMPDLP  
DISVFNLOVIRIRILHNGAYSLTLOGIGISMLGRSLRELGSGLALIHNTLHCFVHTVPMQDLFRNPH  
QALLHTANRPEDECVGEGLAGHOLCARGHCMGPGPTQCVNCSQFLRGOECVEECRVLOGLPREYVNAHCL  
LCHPECOBONGSVTCFPGPADQVCAHAKYKDPFCVARCPSGVKPDLSYMPIMKFPDEBGAQCPPTNC  
THSCVDLDKGCFAEORASPLTISIVSAVVGILLVVLGVFGILIKRQOKIRKTYTRLLQETELVEPL  
TSGAMPNOQMRILKETELRKXKVLGSAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDE  
AYMAGVGSPPVYSRLIGICTSTVOLVTOQLMPYGCILDHVRNENGRGLGSODLNMWQIAKMSYLEDR  
LVHRLDAARNVLYKSPNHVKITDGLARLDDIDETEXHADGKVPIMKMALESILRRRFTHOSDYMAYV  
TYWELTFGAKPYDGI PAREIPDLLEKGERLPPOPICITIDVYMIWVKCMIMDSECRPRFRELVS  
RPORFVVIIONEDIGASPLDSTFYRSILLEDDMDGLVDAEYLVPOGCFPCDPAPAGAGMHHRRSS  
STRSGGDLTLGLEPSEEBEAPRSLAPSEBAGSDVFDGDLGMKAAGLQSLPTHDPSPLORYSEDPVPL  
PSTIDGYVAPLTCSPQPEYVNOQDVRRPOPSPREGPLPAARPAAGATLERAKTLSPGNQGVVKQVFAFGA  
VENPEYLTPOGGAAPQPHPPAPAFSPADNLVYWDODPPERGAPPSTFKGPTAENPEYLGIDVPI

; Entered [jdelaval 22-Jul-03 6:47]  
SEQ4-210-224-14  
MEIAALCRMGILLALILPGAASTOVCTGDMKRLPASPETHLDMRLHLYOGCQVNOGNLELTLYPTNAS  
LSFLQDIOEVQGYVLIJAHNOVROVPLORLIRVGTOLFEDNYALAVLDNNDPLNNTTPTVGTASPGGLREL  
OLRLTEILKGGVLIQNNPOLCYODTILMKDI FHKNNQALTLIDTNRSRACHPCSPMCKGSRMGESSE  
DCOSLTRVCAGGACARCKGFLPFIkanrkfigtelkhsdCLACHFNHSGICEHLCPALVTYNTDTE  
SMNPREGRYTFGASCVTACPYNLSTDVGSCTVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHL  
REVRATYSANIQEFACCKIFGSLAPLPSFDDPASNTPLOPEOLQVETLEIEITGYLYISAMPDLP  
DISVFNLOVIRIRILHNGAYSLTLOGIGISMLGRSLRELGSGLALIHNTLHCFVHTVPMQDLFRNPH  
QALLHTANRPEDECVGEGLAGHOLCARGHCMGPGPTQCVNCSQFLRGOECVEECRVLOGLPREY  
VNAHCLPCHPECOBONGSVTCFPGPADQVCAHAKYKDPFCVARCPSGVKPDLSYMPIMKFPDEBGAQCP  
PCPINCTHSCVDLDKGCFAEORASPLTISIVSAVVGILLVVLGVFGILIKRQOKIRKTYTRLLQET  
ELVEPLTSGAMPNOQMRILKETELRKXKVLGSAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKAN  
KEILDEAYYMAAGVSPVYSRLIGICTSTVOLVTOQLMPYGCILDHVRNENGRGLGSODLNMWQIAKMS  
YLEDRVLVHRLDAARNVLYKSPNHVKITDGLARLDDIDETEXHADGKVPIMKMALESILRRRFTHOSD  
YMSGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICITIDVYMIWVKCMIMDSECRPRFRELVS  
EFSRMARDPQRFVVIIONEDIGASPLDSTFYRSILLEDDMDGLVDAEYLVPOGCFPCDPAPAGAGMHH  
HRRSSSTRSGGDLTLGLEPSEEBEAPRSLAPSEBAGSDVFDGDLGMKAAGLQSLPTHDPSPLORYSE  
DPVPLPSEYDGYVAPLTCSPQPEYVNOQDVRRPOPSPREGPLPAARPAAGATLERAKTLSPGNQGVVKQV  
FAFGAVENPEYLTPOGGAAPQPHPPAPAFSPADNLVYWDODPPERGAPPSTFKGPTAENPEYLGIDVPI  
VI

: Entered [jdelaval 22-Jul-03 6:51]  
SEQ4-250-264-12  
MELALCRWGLLALLPFGAASSTOYCTGTDMKRLRIPASPETHLMDLRHL YQCCQVYQGNLEITYLPTNAS  
LSFLQDIOEVGYVLI IAHNOVROVPLQRLRIVRGTOLEFEDNYALALVLDNGDPLNNTTPTVTGASPGGLREB  
QURSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNOALATLLIDTNRSRACHPCS PMSKGRSCWGESSE  
DCOSLTRIVCAGGACRGKCPPLTDCCHQCAAGCTGPGKSDCLACIHFHNSGICELHCPALVYfn1kansk  
f1g1telRYTFGASCVTACPVNYLSTDVSGCTLVCPILHNOEVTADGTORCEKSCPKARVYGLGMEHL  
REVAVANSANTIOEFAGCKKIFGSLAFIPESFGDDPASNTAPLOPELOVFEETLEETIGYLYISAMPDSL  
DLVAFONLOVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELGSGLALIHNTHLCFVHTVPMDOQLFRNPH  
QALHTANRPEDECEBGLACHQLCARGHCMGPGPTQCVCNSQFLRGQECVEECRVLOGLPREYVNAHCH  
LPCHECOPONGSVTCFGEADQCVACAHYKDPFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPIINC  
THSCVDLDDKCCPAEQRASPLTSIVSAVVGILLVVVLGVFGILLIKRQOKIRKXTMRLLQETELVEPL  
TPSGAMPQAOAMRILKETELRKVKVLSGAGFTVYKGIWIPGENVKIPIVAILKXVLRNTSPKANKELIDE  
AYVAGVSPYVSRILGICLTSTVOLVTOAMPYGCILDHVRENRGLSGODLNMCMQIAKMSYLEDR  
LVHRDLAARNVLVKS PNHVKITDREGARLIDIDEYHADGKVPILKMMALESILRRRTFHQSD  
TVMELMTFGAPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVKCMIDSECRPRFRELVSFSMA  
RDPORFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCDDPAPGAGVHHRHRS  
STRSGGDLTLGLSESEEARSPPLAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPVTPL  
PSETDGVAPLTCSPQPEYVNOVDVRPOPSPRREGPLPAARPAGATLEAKTILSGKNGVNDVAFGGA  
VENBEYLTPOGGAAPQHPHPPAFSFAFDNLVYWDODPPERGAPPSTFKGTPTAENBEYLGLDVPI

: Entered [jdelaval 22-Jul-03 6:51]  
SEQ4-250-264-14  
MELALCRWGLLALLPFGAASSTOYCTGTDMKRLRIPASPETHLMDLRHL YQCCQVYQGNLEITYLPTNAS  
LSFLQDIOEVGYVLI IAHNOVROVPLQRLRIVRGTOLEFEDNYALALVLDNGDPLNNTTPTVTGASPGGLREB  
QURSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNOALATLLIDTNRSRACHPCS PMSKGRSCWGESSE  
DCOSLTRIVCAGGACRGKCPPLTDCCHQCAAGCTGPGKSDCLACIHFHNSGICELHCPALVfn1kansk  
f1g1telRYTFGASCVTACPVNYLSTDVSGCTLVCPILHNOEVTADGTORCEKSCPKARVYGLGMEHL  
REVAVANSANTIOEFAGCKKIFGSLAFIPESFGDDPASNTAPLOPELOVFEETLEETIGYLYISAMPDSL  
DLVAFONLOVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELGSGLALIHNTHLCFVHTVPMDOQLFRNPH  
QALHTANRPEDECEBGLACHQLCARGHCMGPGPTQCVCNSQFLRGQECVEECRVLOGLPREYVNAHCH  
LPCHECOPONGSVTCFGEADQCVACAHYKDPFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPIINC  
THSCVDLDDKCCPAEQRASPLTSIVSAVVGILLVVVLGVFGILLIKRQOKIRKXTMRLLQETELVEPL  
TPSGAMPQAOAMRILKETELRKVKVLSGAGFTVYKGIWIPGENVKIPIVAILKXVLRNTSPKANKELIDE  
AYVAGVSPYVSRILGICLTSTVOLVTOAMPYGCILDHVRENRGLSGODLNMCMQIAKMSYLEDR  
LVHRDLAARNVLVKS PNHVKITDREGARLIDIDEYHADGKVPILKMMALESILRRRTFHQSD  
TVMELMTFGAPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVKCMIDSECRPRFRELVSFSMA  
RDPORFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCDDPAPGAGVHHRHRS  
STRSGGDLTLGLSESEEARSPPLAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPVTPL  
PSETDGVAPLTCSPQPEYVNOVDVRPOPSPRREGPLPAARPAGATLEAKTILSGKNGVNDVAFGGA  
VENBEYLTPOGGAAPQHPHPPAFSFAFDNLVYWDODPPERGAPPSTFKGTPTAENBEYLGLDVPI



; Entered [jdelava] 22-Jul-03 7:10]  
SEQ4-325-339-12  
MEFLALCRMGILLALLPFGAASQVCTGTDMLKRLPASPETHLDMRLHYOGCOVVOGNLELTYLPTNAS  
LSFLADIOEVGVYLIAHNOVROVPLQRLRIVRGTOLFEDNYALAVLDNGDPLNNTTPVTGASRGGLREL  
QLASLTIELKGGVLIQRNPOLCTODTILMKDIFHKNNQALTLIDTNRSAHCPCSPMCKGSRMGESSE  
DCOSLTRTVAGGCARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELCPALVTYNTDTFE  
SMNPBGRVYFGASCYACPYNTLSTDVGCTLVCPPLHNOBVTAEDEGTORCEKSKPCARVCYGLMqy1  
kansk1gltelEFAGCKI1FGSLAFLPESFDGDPASNTAPLOEQLOVEFTLEETGYLYISAMPDPLP  
DLSPFONLOVIRGRILHNGAVSLTLQGLISWGLRSLRELSGLALIHNTHLCFVHTVPMDOLPFNPH  
OALLHTANRPDEBCVGGGLACHOLCARGHCGMGPPTQCVNCSQFLRGOECVBEGRVLOGLPREYVNAHRC  
LPCHPCOPNGSVTCGPEADQCAHAKDPPCVARCPGVPDLSTYMP1WKFPDEBEGACQPCPINC  
THSCVLDLDKGCAPBQASPLTSIVSAVVGILLVVLGVVPGIILKROQKIRKTYMRLLQETELVEPL  
TPSGAMPNQOMRILKETELKXKVLGSGAFGVYKGIWLPDGENVKI1PAIKVLRNTSPKANKELIDE  
AYVMAVGSPYVSRLLGI1CLTSTVOLVTQMLPVGCLDHYENRGRIGSODLNMCMQI1AKMSYLEDR  
LVHRLDAARVNLVKSPPHVKITDFFGLARLLDIDETEYHADGKVP1KMMALESTILRRFTHOSVMSYGV  
TWELMTFGAKPYDGI1PAREIPDLLEKGERLPOPICTIDVYMI1WKCMIDSECRPRPRELTVSEFSRMA  
ROBORFVU1ONEDLGSPASPLDSTFYRSILEDDMGDLVDAEYLVPOOGFCPPAPAGAGMTHHRRSS  
STPSGGGDLTLGLPSEEEAPRSPPLASBEGAGSVFPGDILGMCAKGLQSLPTHDPFLQRYSEDPVPLP  
PSETDGYVAPLTCSPQPEYVNOQDVPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFARFGA  
VENPEYLLTPQGAAPQHPHPAPFSPAFDNLVYWDQDPPERGAAPPSTFKGPTTAENPEYLGIDVPI

; Entered [jdelava] 22-Jul-03 7:10]  
SEQ4-325-339-14  
MEFLALCRMGILLALLPFGAASQVCTGTDMLKRLPASPETHLDMRLHYOGCOVVOGNLELTYLPTNAS  
LSFLADIOEVGVYLIAHNOVROVPLQRLRIVRGTOLFEDNYALAVLDNGDPLNNTTPVTGASRGGLREL  
QLASLTIELKGGVLIQRNPOLCTODTILMKDIFHKNNQALTLIDTNRSAHCPCSPMCKGSRMGESSE  
DCOSLTRTVAGGCARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELCPALVTYNTDTFE  
SMNPBGRVYFGASCYACPYNTLSTDVGCTLVCPPLHNOBVTAEDEGTORCEKSKPCARVCYGLMq1n  
ftvafw1rydkvaasnlEEFAGCKI1FGSLAFLPESFDGDPASNTAPLOEQLOVEFTLEETGYLYISA  
MPDLPDLSPFONLOVIRGRILHNGAVSLTLQGLISWGLRSLRELSGLALIHNTHLCFVHTVPMDO  
LAFNPQALLHTANRPDEBCVGGGLACHOLCARGHCGMGPPTQCVNCSQFLRGOECVBEGRVLOGLPREY  
VNAHRC1PCHPECOPONGSVTCGPEADQCAHAKDPPCVARCPGVPDLSTYMP1WKFPDEBEGACQ  
PCP1NCTHSCVDLDKGCAPBQASPLTSIVSAVVGILLVVLGVVPGIILKROQKIRKTYMRLLQET  
ELVEPLTPSGAMPNQOMRILKETELKXKVLGSGAFGVYKGIWLPDGENVKI1PAIKVLRNTSPKAN  
KEILDEAYVMAVGSPYVSRLLGI1CLTSTVOLVTQMLPVGCLDHYENRGRIGSODLNMCMQI1AKMS  
YLEDRVLVHRLDAARVNLVKSPPHVKITDFFGLARLLDIDETEYHADGKVP1KMMALESTILRRFTHOS  
VMSGYVTWELMTFGAKPYDGI1PAREIPDLLEKGERLPOPICTIDVYMI1WKCMIDSECRPRPRELTVS  
EFSHMAEDPQRFVU1ONEDLGSPASPLDSTFYRSILEDDMGDLVDAEYLVPOOGFCPPAPAGAGMTH  
HRRHSSSTRSGGDLTLGLPSEEEAPRSPPLASBEGAGSVFPGDILGMCAKGLQSLPTHDPFLQRYSE  
DPVPLPSETDGYVAPLTCSPQPEYVNOQDVPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKD  
VFAFGAVENPEYLLTPQGAAPQHPHPAPFSPAFDNLVYWDQDPPERGAAPPSTFKGPTTAENPEYLGIDVP  
VI

! Entered [jdela] 22-Jul-03 7:14]  
SEQ4-369-383-12  
MEIAALCRMGULALALPRGAASOVCTGTMKLR.LPASPTHLDMLRHL.YOGCQVOGNNLE.TYLP.TNAS  
LSFLQDIQEVQGYVLI.AHNQVROVPLQRL.IVRGTQUL.FEDNYALAVLDNGDPLNNTT.PVTGASFGJREL  
QLRSLTEILKGVLI.QRNPOLCYODTILMKDI.FHKNNQALATLIDITNRSRACHPCSPMCKGSRMGESSE  
DCOSLRTVYAGGACGACRKGPL.PTDCHEOCAGCTGPKHSDCLALHFNHSGICELHC.PALVYNTDTFE  
SMNPEGRYTFGASCVTACPVNYLSTVGSCTIVCP.LHNOEVTABDGTORCEKSKPCARVCYGLGMqYI  
kanakfi.gite.l.e.fEAGCKI.FGSLAFL.PESFDGDPASNTA.PqYI.kanakfi.gite.l.tGYLYISAMPDILP  
DLSVPONLQVIRGRI.LHNGAYSLTLQGLISWLGRLSRELGSGLAL.IHNTHL.CFVHTYPMQDL.FRNPH  
QALLHTANRPEDCEVGEGLACHQUL.CARGHCKMGPPTOCVNC.SQFLRGOECVEBCRVLOGL.FREYVNAHRC  
LPHPBCQOPONGSVTCF.GPEADQCVACAHYKDPFCVARCPSGVK.PDLSTYMPIMKFPDEBACQPCINC  
TSCVLDLDDKCGCPAEORASPLT.SIVSAVVGILV.VLVGVVFGILIKRQOKIRKYTMRRLI.QOETELVEPL  
TPSGAMPNQAQMRILKETELRKXVVLGSGAFVYKGIWI.PDGENVKI.PVAIKVLRNTSPKANKEI.LDE  
AYMAGVGSYVSRLIGICL.TSTYQVLTQ.LMPYGC.LLDHRENRGR.LGSODL.LWCMQI.AKMSYLEDVR  
LVHRDLAARVVLV.KSPHVKITDGLAR.LLDIDETEYHADGKVPIKMMALESTLRRTFHOSDVMSTYV  
TYWELMTFGAKPYDGI.PAREIPDLLEKGERL.POPICTIDVYMI.VKCMIMIDSECRPFRELVSEFSRMA  
RDQRFVVI.QNEDLGPASPLDSTFYRS.LLEDDBDMDLVDAEYLV.PQOGFPCPDPAAGAGVMHHRSS  
STRSGGDDLTLGLEPSEEARSP.LAPSEBAGSDVPDGLGMGA.KGLQSL.PTHDPSPLQRYSEDPTVPL  
PSETDGVVAAL.TCSPOBEYVNO.PVRRPO.PSPREGPL.PAAR.PAGATLERAKT.LSPGXGVVXKDVAFAGGA  
VENPEY.LTPQGAAPQHP.PPAPSPAFDNL.YYMDQDPERGAPSTFKGPTAENPEYLGLDVPI

! Entered [jdela] 22-Jul-03 7:14]  
SEQ4-369-383-14  
MEIAALCRMGULALALPRGAASOVCTGTMKLR.LPASPTHLDMLRHL.YOGCQVOGNNLE.TYLP.TNAS  
LSFLQDIQEVQGYVLI.AHNQVROVPLQRL.IVRGTQUL.FEDNYALAVLDNGDPLNNTT.PVTGASFGJREL  
QLRSLTEILKGVLI.QRNPOLCYODTILMKDI.FHKNNQALATLIDITNRSRACHPCSPMCKGSRMGESSE  
DCOSLRTVYAGGACGACRKGPL.PTDCHEOCAGCTGPKHSDCLALHFNHSGICELHC.PALVYNTDTFE  
SMNPEGRYTFGASCVTACPVNYLSTVGSCTIVCP.LHNOEVTABDGTORCEKSKPCARVCYGLGMqYI  
kanakfi.gite.l.e.fEAGCKI.FGSLAFL.PESFDGDPASNTA.PqYI.kanakfi.gite.l.tGYLYISAMPDILP  
DLSVPONLQVIRGRI.LHNGAYSLTLQGLISWLGRLSRELGSGLAL.IHNTHL.CFVHTYPMQDL.FRNPH  
QALLHTANRPEDCEVGEGLACHQUL.CARGHCKMGPPTOCVNC.SQFLRGOECVEBCRVLOGL.FREYVNAHRC  
LPHPBCQOPONGSVTCF.GPEADQCVACAHYKDPFCVARCPSGVK.PDLSTYMPIMKFPDEBACQPCINC  
TSCVLDLDDKCGCPAEORASPLT.SIVSAVVGILV.VLVGVVFGILIKRQOKIRKYTMRRLI.QOETELVEPL  
TPSGAMPNQAQMRILKETELRKXVVLGSGAFVYKGIWI.PDGENVKI.PVAIKVLRNTSPKANKEI.LDE  
AYMAGVGSYVSRLIGICL.TSTYQVLTQ.LMPYGC.LLDHRENRGR.LGSODL.LWCMQI.AKMSYLEDVR  
LVHRDLAARVVLV.KSPHVKITDGLAR.LLDIDETEYHADGKVPIKMMALESTLRRTFHOSDVMSTYV  
TYWELMTFGAKPYDGI.PAREIPDLLEKGERL.POPICTIDVYMI.VKCMIMIDSECRPFRELVSEFSRMA  
RDQRFVVI.QNEDLGPASPLDSTFYRS.LLEDDBDMDLVDAEYLV.PQOGFPCPDPAAGAGVMHHRSS  
STRSGGDDLTLGLEPSEEARSP.LAPSEBAGSDVPDGLGMGA.KGLQSL.PTHDPSPLQRYSEDPTVPL  
PSETDGVVAAL.TCSPOBEYVNO.PVRRPO.PSPREGPL.PAAR.PAGATLERAKT.LSPGXGVVXKDVAFAGGA  
VENPEY.LTPQGAAPQHP.PPAPSPAFDNL.YYMDQDPERGAPSTFKGPTAENPEYLGLDVPI  
VI

; Entered [jdelaval 22-Jul-03 7:17]  
SEQ4-465-479-12  
MELALACRMWGLLALPPGASTOVCTGTDKMLRLPASPTHLDMLRHLVGGCCVAVGQNLLETTLPTNAS  
LSPLDIOEVGGVLIANNOVROVPLORLIRVGTOLPEDNYALAVLNGDPLNNTTPTVTSRPGSLREL  
OLRSTTEILKGGVLIORNPOLCYODTILMKDI FHKNNOLALTIDTNSRACHPCSPMKGSRCKGSESE  
DCOSLTRIVCAGGACRCRCKPLPTDCCHEQCAAGCTGPRHSDCLACLHNHSGICEHLCPALVTYNTDTFE  
SMNPEGRYTFGASCCTACCPYNYLSTDVGSCTLVCPLNHÖEVTABDGTÖRCEKSKPCARVCYGLMGYI  
kanakfigiteie1EPAGCKI1FGSLAFLPESFDGPASNTAPLOPEQLQVFETLBEITGYLYISAMPDLP  
DLVSFONLOVIRGRILHNHGAISLTLOGLISMLGRSLRELSGIALIHNHTHLCFVHTVMDOLFRRGYI  
kanakfigiteie1ECVGGIACHQICARGHCWGPÖTQCVNCSQFLRGOECVECECVLQGLPREYVNAHRC  
LPCHECOPÖNGSVTCRGPÖADQVACAHYKDPPECVARCPSGVKPDLSYMPIMKFPDEBAGCOPCINC  
THSCVDLDKGCPRAPÖRASPLTSIVSAVVGILLVVILGVFGILLIKRQÖKIRKTYMRLLQETELVEPL  
TPSGAMNÖKÖMRLKETELRKVYLSGAGFTYKGIWIPDGENVKI1PAIKVIRENTSPRANKELIDE  
AYVAGVGSPIVSRLLGICLTSTVQVLTÖLMPYGCGLDHWRENRGRILGSÖDLLNMCQI1AKGMSYLEDR  
LVHRDLAARNVLVSPHNVKITDQGLARLDDIDETEHADGKVPI1KMMALESILRRRFTHQS  
TVMLMTFGAKPYDGI1PAREIPDLLEKGERLPOPPICTIDYMIWVKCMIDSECRPRELVSSEFMA  
RDPÖRFVVIÖNEDLSPASPLDSTFYRSILLEDDMDGLVDAEYLVPOQGFPCDPDPAAGAGMHHRRSS  
STRSGGDLTLGLEPSEEAARSPILASBEGAGSDVFDGLMGAKGLQSLPTHDPSPLÖRYS  
PSETDGVAPLITCSPOPEYVNOÖVROPSPSREGPLPAARPAÖATLERAKTILSPGKGVKDVFAFGGA  
VENPEYLTPOGGAAPQPHPPAFSPAFDNLVYWDÖDPPERGAPSTFKGTPTAENPEYLGIDVPI

; Entered [jdelaval 22-Jul-03 7:17]  
SEQ4-465-479-14  
MELALACRMWGLLALPPGASTOVCTGTDKMLRLPASPTHLDMLRHLVGGCCVAVGQNLLETTLPTNAS  
LSPLDIOEVGGVLIANNOVROVPLORLIRVGTOLPEDNYALAVLNGDPLNNTTPTVTSRPGSLREL  
OLRSTTEILKGGVLIORNPOLCYODTILMKDI FHKNNOLALTIDTNSRACHPCSPMKGSRCKGSESE  
DCOSLTRIVCAGGACRCRCKPLPTDCCHEQCAAGCTGPRHSDCLACLHNHSGICEHLCPALVTYNTDTFE  
SMNPEGRYTFGASCCTACCPYNYLSTDVGSCTLVCPLNHÖEVTABDGTÖRCEKSKPCARVCYGLMGYI  
kanakfigiteie1EPAGCKI1FGSLAFLPESFDGPASNTAPLOPEQLQVFETLBEITGYLYISAMPDLP  
DLVSFONLOVIRGRILHNHGAISLTLOGLISMLGRSLRELSGIALIHNHTHLCFVHTVMDOLFRRGYI  
kanakfigiteie1ECVGGIACHQICARGHCWGPÖTQCVNCSQFLRGOECVECECVLQGLPREY  
fvefwi1vpevvaah1aECVGGIACHQICARGHCWGPÖTQCVNCSQFLRGOECVECECVLQGLPREY  
VNAHRCRPCHECOPÖNGSVTCRGPÖADQVACAHYKDPPECVARCPSGVKPDLSYMPIMKFPDEBAGC  
PCPINCHSCVDLDKGCPRAPÖRASPLTSIVSAVVGILLVVILGVFGILLIKRQÖKIRKTYMRLLQET  
ELVEPLTPSGAMNÖKÖMRLKETELRKVYLSGAGFTYKGIWIPDGENVKI1PAIKVIRENTSPRANKELIDE  
AYVAGVGSPIVSRLLGICLTSTVQVLTÖLMPYGCGLDHWRENRGRILGSÖDLLNMCQI1AKGMSYLEDR  
LVHRDLAARNVLVSPHNVKITDQGLARLDDIDETEHADGKVPI1KMMALESILRRRFTHQS  
TVMLMTFGAKPYDGI1PAREIPDLLEKGERLPOPPICTIDYMIWVKCMIDSECRPRELVSSEFMA  
RDPÖRFVVIÖNEDLSPASPLDSTFYRSILLEDDMDGLVDAEYLVPOQGFPCDPDPAAGAGMHHRRSS  
STRSGGDLTLGLEPSEEAARSPILASBEGAGSDVFDGLMGAKGLQSLPTHDPSPLÖRYS  
PSETDGVAPLITCSPOPEYVNOÖVROPSPSREGPLPAARPAÖATLERAKTILSPGKGVKDVFAFGGA  
VENPEYLTPOGGAAPQPHPPAFSPAFDNLVYWDÖDPPERGAPSTFKGTPTAENPEYLGIDVPI  
V1

: Entered [jdelaval 22-Jul-03 7:20]  
SEQ4-579-593-12  
MELALACRWSGLLALLPRGAASVCTGTDMKRLRPAPEETHLDMRLHYGCCQVYQGNLEITYLPTNAS  
LSFLDODIEVGGVLIHANVRQVPLQRLRIVRGTLFEDNYALAVLNDGDLNNTTVPVTGASPGGLRELB  
QRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLALTLLIDTNRSRACHPCSPMGKGRCSGSESE  
DCOSLTRIVCAGGACARCKGPLPTDCHEQCAAGCTGPKHSDCLACLHFHNSGICELHCPALVTYNTDPE  
SMNPBEGYTTGASCVTACPYNYLSTDVSGTLVCPILHNOEYTAEDGTORCEKSGPCARVCYGMGYI  
kanefiigitelEFGACKKIFGSLAFIPESFDGDPASNTAPLOPQLOVFEETLEITGYLYISAMPDILP  
DLVFNQIVIRGRILHNGAVSLTQIGIGISWLGRLSRLRELSGLALIHNNHLCVHTVPMDOULFRNPH  
QALLHTANRPEDECVBEGGLACHOLCARHCWGPPTQCVCNSQFLRGQECVEECRVLOGLPREYVNAHRC  
LPCHRECOPONGSVTCFGEADQVCAHAHYKDPFCVACPGYIkanefiigitelPDEGACQCPPIHC  
THSCVDLDDKCGCPAQRASPLTSIVSAVVGILLVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPL  
TPSGAMPNQAQMRILKETELRKVYLGSAGFTVYKGIWIPGENVKIPVAIKVIRENTSPRANKEILDE  
AYMAGVSGPYVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLSGSODLLNMCMOIAKMSYLEDVR  
LVHRDLAARNVLVKSBNHVKITDPGLARLLDIDETEHADGKVPKMMALESIILRRFTTHOSDWSYCV  
TWELMTGAKPYDGI PARELIDLEKGERLPQPICTIDVTMIMVKCMWIDSECPREFELVSEFSRA  
RDPORFVVIQNEIDLGPASPLDSTFYRSLLEDMDMDLVDAEYLVPOQGFCDPAPAGAGVHHRRSS  
STRSGGDLTLGLPSEEBEAPRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDPVVL  
PSETDGYVAPLTGCSQPEYVNOQDVROPSPREGPLPARPAGATLEBAKTLISPKNGVVDVAFGGA  
VENPEYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVPI

: Entered [jdelaval 22-Jul-03 7:20]  
SEQ4-579-593-14  
MELALACRWSGLLALLPRGAASVCTGTDMKRLRPAPEETHLDMRLHYGCCQVYQGNLEITYLPTNAS  
LSFLDODIEVGGVLIHANVRQVPLQRLRIVRGTLFEDNYALAVLNDGDLNNTTVPVTGASPGGLRELB  
QRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLALTLLIDTNRSRACHPCSPMGKGRCSGSESE  
DCOSLTRIVCAGGACARCKGPLPTDCHEQCAAGCTGPKHSDCLACLHFHNSGICELHCPALVTYNTDPE  
SMNPBEGYTTGASCVTACPYNYLSTDVSGTLVCPILHNOEYTAEDGTORCEKSGPCARVCYGMGYI  
kanefiigitelEFGACKKIFGSLAFIPESFDGDPASNTAPLOPQLOVFEETLEITGYLYISAMPDILP  
DLVFNQIVIRGRILHNGAVSLTQIGIGISWLGRLSRLRELSGLALIHNNHLCVHTVPMDOULFRNPH  
QALLHTANRPEDECVBEGGLACHOLCARHCWGPPTQCVCNSQFLRGQECVEECRVLOGLPREYVNAHRC  
LPCHRECOPONGSVTCFGEADQVCAHAHYKDPFCVACPGYIkanefiigitelPDEGACQCPPIHC  
THSCVDLDDKCGCPAQRASPLTSIVSAVVGILLVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPL  
TPSGAMPNQAQMRILKETELRKVYLGSAGFTVYKGIWIPGENVKIPVAIKVIRENTSPRAN  
KEILDEAYMAGVSGPYVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLSGSODLLNMCMOIAKMS  
YLEDVRLVHRDLAARNVLVKSBNHVKITDPGLARLLDIDETEHADGKVPKMMALESIILRRFTTHOSD  
WVSYGVTWELMTGAKPYDGI PARELIDLEKGERLPQPICTIDVTMIMVKCMWIDSECPREFELVSE  
EFSRNARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDMDMDLVDAEYLVPOQGFCDPAPAGAGVHH  
HHRSSSTRSGGDLTLGLPSEEBEAPRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSE  
DPTVLPSETDGYVAPLTGCSQPEYVNOQDVROPSPREGPLPARPAGATLEBAKTLISPKNGVVDVAF  
FAGGAVENPEYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVPI  
VI

; Entered [jdelaval 22-Jul-03 7:26]  
SEQ4-632-652-12  
MELAAICRMGLLALLPBGAASTOVCTGTMKRLPASPETHLDMLEHLXOGCOVQGNLELTYPNTNAS  
LSFLDIOIEVOGYVLIHNOVROVPLORLIRVGTOLPEDNYALAVLDNGDPLNNTTPVTGASRGGLREL  
QARSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNQALTLIDTRSRACHPCSPWCKGSRCMGESSE  
DCOSLTRIVCAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFFE  
SMNPEGRYTFGASCYACVYNYLSTDVGSCTLVCPPLHNOBVTAEADGTORCEKSKPCARVCYGLMGYI  
kanakfigiteLEFAGCKKIFGSLAFIPESFDDGPASNTAPLOEOLQVEFTELEITGYLISAMPDILP  
DLSEFONLQVIRGRILHNGAYSLTLOGLISMLGRSLRELGLALIHNTHLCPVHTVPMWDLFRNPH  
QALLHTANRPEDECVGEGLACHOLCARHCWGPGPTQVCNCSOFLRGOECVEBCRVLOGLPREVYNAHRC  
LCHPECQPNQNGSVTCFGEPADQVACAHYKDPFCVACRPSGVKPDLSYMP1WKFPDEEGACQPCPINC  
THSCVDLDKNGCAEORASPLTSI1qy1kanakfigiteLKRQOKIRKYTMRLLOETELVEPLTPSGAM  
PNOQMRILKETELRKVYVLSGAFGVYKGIWIPDENVKIPVAIKVLRENTSPKANKEILDEAVYVAG  
VCSYVSRLLIGICTSTVOLVTOIMPYGCILDHYRENRGRIGSODLNMCMQIAKMSYLEDEVR  
AARNVYVKS PNHVKITDQGLARLLDIDETEHADGKVPIKMMALIESILRRRFTHOSDWSYGVTVWELM  
TVGAKYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVCMWIDSECRPRELIVSEFSRMA  
VVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDAEYILVPOQGFCCPDPAPAGAGVHHHRSSSTRSGG  
GDLTLGLEPSEEEAPRSPLAPSEGASDVFDGDLGMAAKGLOS1PTHDPSPLOQRYSEDPVPLPSESTDG  
YVAPLTCSPQPEYVNOPDVAPQPPSPREGPLPARPAGATLERAKTILSPGKNGVYKDVAFGCAVENPEY  
LTPQGGAAPQHPHPPAFSPAFDNL1YWDQDPPERGAAPSTFKGPTAENPEYLGIDVPV1

; Entered [jdelaval 22-Jul-03 7:26]  
SEQ4-632-652-14  
MELAAICRMGLLALLPBGAASTOVCTGTMKRLPASPETHLDMLEHLXOGCOVQGNLELTYPNTNAS  
LSFLDIOIEVOGYVLIHNOVROVPLORLIRVGTOLPEDNYALAVLDNGDPLNNTTPVTGASRGGLREL  
QARSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNQALTLIDTRSRACHPCSPWCKGSRCMGESSE  
DCOSLTRIVCAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFFE  
SMNPEGRYTFGASCYACVYNYLSTDVGSCTLVCPPLHNOBVTAEADGTORCEKSKPCARVCYGLMGYI  
kanakfigiteLEFAGCKKIFGSLAFIPESFDDGPASNTAPLOEOLQVEFTELEITGYLISAMPDILP  
DLSEFONLQVIRGRILHNGAYSLTLOGLISMLGRSLRELGLALIHNTHLCPVHTVPMWDLFRNPH  
QALLHTANRPEDECVGEGLACHOLCARHCWGPGPTQVCNCSOFLRGOECVEBCRVLOGLPREVYNAHRC  
LCHPECQPNQNGSVTCFGEPADQVACAHYKDPFCVACRPSGVKPDLSYMP1WKFPDEEGACQPCPINC  
THSCVDLDKNGCAEORASPLTSI1fnf1vsfw1ydkvaashLEKROOKIRKYTMRLLOETELVEPL  
TPSGAMPNOQMRILKETELRKVYVLSGAFGVYKGIWIPDENVKIPVAIKVLRENTSPKANKEILDE  
AYVAGVSPYSRLLIGICTSTVOLVTOIMPYGCILDHYRENRGRIGSODLNMCMQIAKMSYLEDEVR  
LVHRDLAARNVYVKS PNHVKITDQGLARLLDIDETEHADGKVPIKMMALIESILRRRFTHOSDWSYGV  
TVWELMTFGAKPTDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVCMWIDSECRPRELIVSEFSMA  
RDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDAEYILVPOQGFCCPDPAPAGAGVHHHRSS  
STRSGGDLTLGLEPSEEEAPRSPLAPSEGASDVFDGDLGMAAKGLOS1PTHDPSPLOQRYSEDPVPL  
PSETDGVVAPLTCSPQPEYVNOPDVAPQPPSPREGPLPARPAGATLERAKTILSPGKNGVYKDVAFGCA  
VENPEYLTPOGGAAPQHPHPPAFSPAFDNL1YWDQDPPERGAAPSTFKGPTAENPEYLGIDVPV1

; Entered [jdelaval 22-Jul-03 7:29]  
SE04-653-675-12  
MELALCRWGLLALLPFGAASVCTGTDMKRLRPAPETHLDMRLHYGCCVVOGNTLTYLPTNAS  
LSFLQDIOEVQGYVLIANOVROVPLQRLRIVRGTOLEFEDNYALALVLDNGDPLNNTTPTVTGASPGGLREB  
QRLSLEILKGGVLIQRNPOLCYODTILMKDI FHNNOALALTIDITNRSRACHPCS PMCKSGRCWGESSE  
DCOSLTRVACAGGACRCAGPLPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFE  
SMNPBERYTFGASCVTACPNYVLSITDVSGCTLVCP LHNQEVTAEDGTORCEKSCPKCARVCYGLMGYI  
kanek f i g i t e l e f a g c k k i f g s l a f l p e s f d g d p a s n t a p l o p e l o v f e t l e e i t g y l y i s a m p d s l p  
d l s v f o n i o v r g r i l n h g a v s l t l o g i g i s m l g r s l r e l g s g l a l i h n t h l c f v h t v p m d o l f r n p h  
OALHTANRPEDECVBEGELACHOLCARGHCWGPGTQCVNCSQFLRGQECVEECVLOGLPREYVNAHRC  
LPCHPECOPONGSVTCFGEPEADQVACAHYKDPPTCVARCPBGVAPDLSTYMPIMKFPDEBGAQCPINIC  
THSCVDLDDKGCPEAOBRASPLTISIVSAVVGILLVVLGVFGILIQYI kanek f i g i t e l e l p l t p s g a m p n  
QAOHRLIKETELRKVKVLGSAFGTVYKGIWIPDENVKI PVAIKVLRNTSPKANKKEILDEAYMAGV  
SPYVSRILGICLTSTVOLVTOIMPYGCLLDHVRNRRGLSQDILNMCQI AKGMSYLEVRLVHRDLAA  
RNVLVKSPNHYKITDIFGLARLIDIDEYHADGGKVP I K M M A L E S I L R R R F T H O S D V M S Y G V T W E L M T F  
GAKYDGI PAREI PDLLEKGERLPOPICTIDVYIMVKCMWIDSECRPRFRELVSERSMARDPQRTV  
IONEDLGPASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMWHHRSSSTRSGGD  
LTLEBESBEAPRSPLASSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSBEDPTVPLPSETDGYV  
APLTCSPQEVYNQPDVRRPQPSPREGPLPAARPAAGATLERAKTILSPGNGVVKDVFAGGAVENPEYLT  
POGGAAPQHPHPAFSPAFDNLXYWDQDPPERGAPRSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 7:29]  
SE04-653-675-14  
MELALCRWGLLALLPFGAASVCTGTDMKRLRPAPETHLDMRLHYGCCVVOGNTLTYLPTNAS  
LSFLQDIOEVQGYVLIANOVROVPLQRLRIVRGTOLEFEDNYALALVLDNGDPLNNTTPTVTGASPGGLREB  
QRLSLEILKGGVLIQRNPOLCYODTILMKDI FHNNOALALTIDITNRSRACHPCS PMCKSGRCWGESSE  
DCOSLTRVACAGGACRCAGPLPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFE  
SMNPBERYTFGASCVTACPNYVLSITDVSGCTLVCP LHNQEVTAEDGTORCEKSCPKCARVCYGLMGYI  
kanek f i g i t e l e f a g c k k i f g s l a f l p e s f d g d p a s n t a p l o p e l o v f e t l e e i t g y l y i s a m p d s l p  
d l s v f o n i o v r g r i l n h g a v s l t l o g i g i s m l g r s l r e l g s g l a l i h n t h l c f v h t v p m d o l f r n p h  
OALHTANRPEDECVBEGELACHOLCARGHCWGPGTQCVNCSQFLRGQECVEECVLOGLPREYVNAHRC  
LPCHPECOPONGSVTCFGEPEADQVACAHYKDPPTCVARCPBGVAPDLSTYMPIMKFPDEBGAQCPINIC  
THSCVDLDDKGCPEAOBRASPLTISIVSAVVGILLVVLGVFGILIFNFTVAFWIRPKVASASHLEPLP  
SGAMNOQOMRILIKETELRKVKVLGSAFGTVYKGIWIPDENVKI PVAIKVLRNTSPKANKKEILDEAY  
MAGVSPYVSRILGICLTSTVOLVTOIMPYGCLLDHVRNRRGLSQDILNMCQI AKGMSYLEVRLV  
HRDLAARNVLVKS PNHYKITDIFGLARLIDIDEYHADGGKVP I K M M A L E S I L R R R F T H O S D V M S Y G V T W E L M T F  
WELMTGAKPYDGI PAREI PDLLEKGERLPOPICTIDVYIMVKCMWIDSECRPRFRELVSERSMARDP  
QRTVIONEDLGPASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMWHHRSSSTR  
RSGGDLTLGLEPSEBEAPRSPLASSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSBEDPTVPLPS  
ETDGVVAPLTCSPQEVYNQPDVRRPQPSPREGPLPAARPAAGATLERAKTILSPGNGVVKDVFAGGAVE  
NPEYLTPOGGAAPQHPHPAFSPAFDNLXYWDQDPPERGAPRSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 7:33]  
SEQ4-695-709-12  
MELALCRWGLLALLPFGAASVCTGTGDMKRLPASPETHLDMRLHLYOGCOVVOGNIELTYLPTNAS  
LSFLQDIOEVGVYLIANOVROVPLQRLRIVRGTOLFEDNYALAVLNDGDLNNTTPTYGASPGGLREB  
QURSLTEILKGVLIQRNPOLCYODTILMKDIFHKNQOLATLITDNRBRACHPGSPMCKGSRGSESE  
DCOSLRTVCGAGCARCKGPLPTDCHCOCAAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTE  
SMNPREGRYTFGASCVTACPNYVLSLTDVGSCTLVCPLNHOEYTAEDGTORCEKSPCARVCYGLMGY1  
kanakf1gite1EFAGCKKIFGSLAFLESPFDGDASNAPLQPEQLQVFETLEETGYLYISAMPDSL  
DLVSFQNLQVIRGRILHNHGAVALTLOGIGISWLGRLSRLRELGSLALIHNNHLCFVHTVPMDOULFRNPH  
QALHTANRPREDECEVGBGLACHOLCARHGWCPGPTOCVNCOSQPLRGOECVECECVLOGLPREYVNAHRC  
LPCHPECOPONGSVTCFRPEADOCVACHYKDPFCVARGSPGVKPDLSYMPIMKRPDEBAGCOPCPINC  
THSCVDLDDKCGPAEQRASPLTSIVSAVAGLILVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPL  
TPSGAMRQAOQMRILKEQYIKANAKF1gite1TYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDE  
AYVAVGSPYVSRLLGICLITSVOLVTQLMPEYGCLLDHVRENRGLSGQDLNMCQIAKGMSYLEDV  
LVHRDLAARNVLVSPNHVKITDQGLARLLIDETEHADGKVP1KMMALESILRRFTHOSDVSXGV  
TWMLMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWKCMMIDSECRPRFRELVSSEFMA  
RDPQRFVVIQNEDEGPASPLDSTFYRSLLEDMDMDLVADEEYLVPOQGFCDPAAPGAGGWVHHRHS  
STRGGGDLTLGLPSEEEAPRSPLAPSEGAGSDVFGDGLGMAAKGLQSLPTHDPSPLOQSRSEDPYPL  
PSETDGYVAPLTCSPQPEYVNOVDVRRPQPSPREGPLPAARPAAGATLERAKTILSPGKGVADVAFGGA  
VENPEYLTPOGGAAPQHPHPAFSPAFDNLYYWDQDPPERGAPPSTFKGPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 7:33]  
SEQ4-695-709-14  
MELALCRWGLLALLPFGAASVCTGTGDMKRLPASPETHLDMRLHLYOGCOVVOGNIELTYLPTNAS  
LSFLQDIOEVGVYLIANOVROVPLQRLRIVRGTOLFEDNYALAVLNDGDLNNTTPTYGASPGGLREB  
QURSLTEILKGVLIQRNPOLCYODTILMKDIFHKNQOLATLITDNRBRACHPGSPMCKGSRGSESE  
DCOSLRTVCGAGCARCKGPLPTDCHCOCAAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTE  
SMNPREGRYTFGASCVTACPNYVLSLTDVGSCTLVCPLNHOEYTAEDGTORCEKSPCARVCYGLMGY1  
kanakf1gite1EFAGCKKIFGSLAFLESPFDGDASNAPLQPEQLQVFETLEETGYLYISAMPDSL  
DLVSFQNLQVIRGRILHNHGAVALTLOGIGISWLGRLSRLRELGSLALIHNNHLCFVHTVPMDOULFRNPH  
QALHTANRPREDECEVGBGLACHOLCARHGWCPGPTOCVNCOSQPLRGOECVECECVLOGLPREYVNAHRC  
LPCHPECOPONGSVTCFRPEADOCVACHYKDPFCVARGSPGVKPDLSYMPIMKRPDEBAGCOPCPINC  
THSCVDLDDKCGPAEQRASPLTSIVSAVAGLILVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPL  
TPSGAMRQAOQMRILKEFNITVSLTWLTVKVAASH1ETVYKGIWIPDGENVKIPVALKVLRENTSPKAN  
KEILDEAYVAVGSPYVSRLLGICLITSVOLVTQLMPEYGCLLDHVRENRGLSGQDLNMCQIAKGMS  
YLEDVRLVHRDLAARNVLVSPNHVKITDQGLARLLIDETEHADGKVP1KMMALESILRRFTHOSD  
VWSYGVTVWMLMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWKCMMIDSECRPRFRELVS  
EESPMARPPQRFVVIQNEDEGPASPLDSTFYRSLLEDMDMDLVADEEYLVPOQGFCDPAAPGAGGWVH  
HHRSSSTRGGGDLTLGLPSEEEAPRSPLAPSEGAGSDVFGDGLGMAAKGLQSLPTHDPSPLOQSRSE  
DPTVLPSETDGYVAPLTCSPQPEYVNOVDVRRPQPSPREGPLPAARPAAGATLERAKTILSPGKGVADV  
PAFGAVENPEYLTPOGGAAPQHPHPAFSPAFDNLYYWDQDPPERGAPPSTFKGPTAENPEYLGLDVPI  
V1

: Entered [jdelaval 22-Jul-03 7:35]  
SEQ4-710-730-12  
MELALCRWGLLLALLPFGAASCTQCTGDMKRLRLPASPEETHLDMRLHLYOGCQVVOGNELELYLPTNAS  
LSFLDIDIEVOGVYLIHNOVQVPLQRLIYRGTOLEFDNYALAVLDNGDPLNNTTPVTGASPGGLREL  
OLRSLTEILKGGVLIQRNPOLCYODTILMKDI.FHKNQALATLIDNRSRACHPCSPMGSGRCSWSESSE  
DCOSLTRVCAGGACARCKGPLETDCHEOCAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTE  
SMNPEGRYTFGASCVTACPVYVLSIDVSGCTLVCPHNOEVTAEEDGTORCEKCSKPCARVCYGLMGVYI  
kanekfigite1EPAGCKKI.FGSLAF.LPES.FGDPASNTAPLOPBOLOVFETLBEITGYLYISAMPDSL  
DLVSFONLOVIRGRILHNGAYS.LTLOGLSI.SWGLRS.LRELSGLALLIHNTHL.CVHTVPMOQLFRNPH  
OALLHTANRPEDECVGEGLACHQOLCARGHCMGPGPTQVCNCSOFLRGDECVBECRVLOGLPREYVNAHRC  
LPCHPECOPONGSVTC.FGPEADQVACAHYKDPFCVACPSGVKPDLSYMPIMKPPDEGACQPCPINC  
THSCVDLDKQGPABORASPLTSIVSAVVGILLVVVLGV.FGILLIKROOKIRKTYMRLLQETELVEPL  
TPSGAMPNOAQMRILKETELRKVYLGSGAFGfny1kanekfigite1VLRENTS.PKANKELIDEAVYMAVG  
VGS.PYVSRLLGICLTSTVO.LTQLMPYGCLLDHVRNRLSGQDLNMCQIAKMSYLEDRVRL  
AARNVLVKS.PNHVKITD.FGLARLLDIDETEHADGKVP.IKMMALESILRRPETHOSDVMSYGVTYWELM  
TFGARPYDGI.PAREIPDLLEKGERLPQPICTIDYVMIWVKCMIDSECRPRERELVSERMAVDPORF  
VVIQNEIDGPPAPLSTFYRSLLEDMDMDLVDAEYLVPOQF.FCPDPA.PGAGGVHHRHSSSTRSGG  
GDLTLGLEPSEEEAPRSPLAPSEAGSDVFDGDLGMAKGLQSL.PTHDPS.PLORYSEDPVPLPSETDG  
VVALPLTCSPOPEYVNO.PDVRPOPSPREBGLPAARPAAGATLERAKTILSGKGVVDVFAFGAVENPEY  
LTPQGAAPQPHPPAFSPAFDNLVYWDQDPPERGA.PSTFKGTPTAENPEYLGIDVPV1

: Entered [jdelaval 22-Jul-03 7:35]  
SEQ4-710-730-14  
MELALCRWGLLLALLPFGAASCTQCTGDMKRLRLPASPEETHLDMRLHLYOGCQVVOGNELELYLPTNAS  
LSFLDIDIEVOGVYLIHNOVQVPLQRLIYRGTOLEFDNYALAVLDNGDPLNNTTPVTGASPGGLREL  
OLRSLTEILKGGVLIQRNPOLCYODTILMKDI.FHKNQALATLIDNRSRACHPCSPMGSGRCSWSESSE  
DCOSLTRVCAGGACARCKGPLETDCHEOCAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTE  
SMNPEGRYTFGASCVTACPVYVLSIDVSGCTLVCPHNOEVTAEEDGTORCEKCSKPCARVCYGLMGVYI  
kanekfigite1EPAGCKKI.FGSLAF.LPES.FGDPASNTAPLOPBOLOVFETLBEITGYLYISAMPDSL  
DLVSFONLOVIRGRILHNGAYS.LTLOGLSI.SWGLRS.LRELSGLALLIHNTHL.CVHTVPMOQLFRNPH  
OALLHTANRPEDECVGEGLACHQOLCARGHCMGPGPTQVCNCSOFLRGDECVBECRVLOGLPREYVNAHRC  
LPCHPECOPONGSVTC.FGPEADQVACAHYKDPFCVACPSGVKPDLSYMPIMKPPDEGACQPCPINC  
THSCVDLDKQGPABORASPLTSIVSAVVGILLVVVLGV.FGILLIKROOKIRKTYMRLLQETELVEPL  
TPSGAMPNOAQMRILKETELRKVYLGSGAFGfntfvsf1.LYOKVSAASH1EVLRENTSPKANKELIDE  
AYMAGVGS.PYVSRLLGICLTSTVO.LTQLMPYGCLLDHVRNRLSGQDLNMCQIAKMSYLEDRVRL  
LTHRDIAARNVLVKS.PNHVKITD.FGLARLLDIDETEHADGKVP.IKMMALESILRRPETHOSDVMSYGV  
TYWELM.TFGAKPYDGI.PAREIPDLLEKGERLPQPICTIDYVMIWVKCMIDSECRPRERELVSERMA  
RDPORFVVIQNEIDGPPAPLSTFYRSLLEDMDMDLVDAEYLVPOQF.FCPDPA.PGAGGVHHRHSS  
STRSGGDLTLGLEPSEEEAPRSPLAPSEAGSDVFDGDLGMAKGLQSL.PTHDPS.PLORYSEDPVPL  
PSETDGYVALPLTCSPOPEYVNO.PDVRPOPSPREBGLPAARPAAGATLERAKTILSGKGVVDVFAFGA  
VENPELUTPQGAAPQPHPPAFSPAFDNLVYWDQDPPERGA.PSTFKGTPTAENPEYLGIDVPV1



```
; Sequence 4, Application US/09806703A
; GENERAL INFORMATION:
; APPLICANT: Steinaa, Lucilla
; APPLICANT: Mouritsen, Soren
; APPLICANT: Gautam, Anand
; APPLICANT: Dalum, Iben
; APPLICANT: Haaning, Jesper
; APPLICANT: Leach, Dana
; APPLICANT: Nielsen, Klaus
; APPLICANT: Karlsson, Gunilla
; APPLICANT: Rasmussen, Peter
; TITLE OF INVENTION: Novel Methods for Therapeutic Vaccination
; FILE REFERENCE: 3631-0109P
; CURRENT APPLICATION NUMBER: US/09/806,703A
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/DK99/00525
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: DK 1998 01261
; PRIOR FILING DATE: 1998-10-05
; PRIOR APPLICATION NUMBER: US 60/105,011
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-806-703A-4
; METALALCRWGLLLALLPGCAASTOVCTGIDMKRLRLPASPETHLMDLRHLVQGGCVVQGNLELYL
; PNASISLFIODIOEVGVVGLIAHNOVROYVPIORLRIVRGTOLPEDNYALAVLNGDPIANTTPTVT
; GASPGRLRLQRLSLTEILKGVLIORNPOLCYQDTILMKDIEHKNQALTLITDRSRACHPC
; SPWCKSGRWGSSSEDCQSLTRIVCAGCARKGKPLPTDCHEQCAAGCTGPHSDCLACLAHFNH
; SGI CELHCPALVTYNTDESMNPGRYTFGASCVTACPCYNLSTDVGSCTLVCPHLNQEVTAE
; DGTORCEKSKPCARVCYGLGMEHLREVAITSANIQEFAGCKIFGSLAFIPESFDGDPASNTA
; PLQPOLQVFEETLEETGLYISAMPDSLPLDSFONLOVIRGRILHNGAYSLTLOGLISWGL
; RSLRELGSGLALIHNNTHLCFVHTVPMOLFRMPHQAHLHTARPEDECVBGLACHOLCARGHC
; MGPGTQCNCQSLRGGQCEVECRVLOGLPREYVNAHRLTCHPRECQPNQSVTCRGPADQCV
; ACAHYKDPFCVACRSGVCPDLSYPMKFPDEBACQPCPINCCHSCVDLDDKCCPAEQRASV
; LITSIVAVVGLLVAVLGVVFGILIKRQOKIRKYIMRRLLOETELVEPLTPSGAMPNQAMRIL
; KETELRKVVLGSGAFVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYVMAVGSP
; YVSRLLGICLTSTVOLVTOMLPYGLDDEYRENRGLGSODLNMCMQIAKMSYLEDVRLVHRD
; LAARVLVSPNHWKITDGLARLIDIDETEVADGQKPIKMALESILRRRFTHOSDVMSYCV
; TYWELMTGAKYIDGIPAREIPDLEKGERLPOPCITIDVYIMYKCMWIDSECRPRELVSE
; FSRMARDFORFVIONEDLGPASPLDSTFYSRLLEDDMDGLVDAEYLVPOQGFPCPDPAAG
; GNVHRRHRSSTRSGGGDLTLGLEPSEEA PRSPLAPSGAGSDVDFGDLGMAAGLQSLPTHD
; PSLQRYSEDPVPLPSETDGVVAPLTCSPOPEYVNPQVRPQPSPREGPLYPARPAGATLERA
; KTLSPKNGCVNVDFAFGAVENPEYLTPOGAAAPQHPPPAPSPAFDNLVYWDQDPERGADPS
; TFKGPTAENPEYLGIDVPVI
```

```
; Sequence 12, Application US/09806703A
; GENERAL INFORMATION:
; APPLICANT: Steinaa, Lucilla
; APPLICANT: Mouritsen, Soren
; APPLICANT: Gautam, Anand
; APPLICANT: Dalum, Iben
; APPLICANT: Haaning, Jesper
; APPLICANT: Leach, Dana
; APPLICANT: Nielsen, Klaus
; APPLICANT: Karlsson, Gunilla
; APPLICANT: Rasmussen, Peter
; TITLE OF INVENTION: Novel Methods for Therapeutic Vaccination
; FILE REFERENCE: 3631-0109P
; CURRENT APPLICATION NUMBER: US/09/806,703A
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/DK99/00525
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: DK 1998 01261
; PRIOR FILING DATE: 1998-10-05
; PRIOR APPLICATION NUMBER: US 60/105,011
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 12
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Clostridium tetani
; US-09-806-703A-12
; QYTKANSKFLGITELI
```

```
; Sequence 14, Application US/09806703A
; GENERAL INFORMATION:
; APPLICANT: Steinaa, Lucilla
; APPLICANT: Mouritsen, Soren
; APPLICANT: Gautam, Anand
; APPLICANT: Dalum, Iben
; APPLICANT: Haaning, Jesper
; APPLICANT: Leach, Dana
; APPLICANT: Nielsen, Klaus
; APPLICANT: Karlsson, Gunilla
; APPLICANT: Rasmussen, Peter
; TITLE OF INVENTION: Novel Methods for Therapeutic Vaccination
; FILE REFERENCE: 3631-0109P
; CURRENT APPLICATION NUMBER: US/09/806,703A
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/DK99/00525
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: DK 1998 01261
; PRIOR FILING DATE: 1998-10-05
; PRIOR APPLICATION NUMBER: US 60/105,011
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO: 14
; LENGTH: 21
; TYPE: PR1
; ORGANISM: Clostridium tetani
US-09-806-703A-14
FNNFTVSPWLRVPKVSASHLE1
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

protein - protein search, using sw model

July 22, 2003, 08:05:49 ; Search time 126.375 Seconds  
(without alignments)  
2046.206 Million cell updates/sec

US-09-806-703A-4

1: 6812  
2: 1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

ence: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 671580 seqs, 206047115 residues

cal number of hits satisfying chosen parameters: 671580

imum DB seq length: 0  
ximum DB seq length: 200000000

st-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

- Database :  
1: SPREMBL\_21: \*  
2: sp\_archaea: \*  
3: sp\_bacteria: \*  
4: sp\_fungi: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriaph: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6295	92.4	1259	6 018735	018735 canis fam1
2	3171	46.6	1209	11 090X70	090X70 rattus norv
3	3142	46.1	1210	11 09EP98	09EP98 mus musculu
4	2751	40.4	1165	13 09VH40	09VH40 xiphophorus
5	2729.5	40.1	1137	13 09K6F6	09K6F6 gallus gall
6	2326	34.1	1328	13 P79754	P79754 fugu rubrip
7	2044.5	30.0	1433	5 09BIH9	09BIH9 anopheles g
8	1874.5	27.5	419	4 09UK79	09UK79 homo sapien
9	1739	25.2	729	11 08R2X1	08R2X1 mus musculu
10	1720	25.2	729	15 086712	086712 avian rous-
11	1697.5	24.9	412	4 08WTV0	08WTV0 homo sapien
12	1653.5	24.3	962	15 064895	064895 avian eryth
13	1645	24.1	545	15 085468	085468 avian eryth
14	1533.5	22.5	655	11 09WVFS	09WVFS mus musculu
15	1517.5	22.3	643	11 09ERV6	09ERV6 mus musculu

17	1285	18.9	1193	5 09Y1X8	09Y1X8 ephydaria f
18	1200.5	17.6	1368	5 023821	023821 caenorhabdi
19	1177	17.3	1717	5 026566	026566 schistosoma
20	1155	17.0	527	13 090836	090836 gallus gall
21	1028.5	15.1	478	11 09SE80	09SE80 rattus norv
22	971.5	14.3	599	13 09PSH2	09PSH2 gallus gall
23	906	13.3	165	4 014256	014256 homo sapien
24	887	13.0	176	11 0923V5	0923V5 rattus norv
25	806.5	11.8	346	13 P11776	P11776 xiphophorus
26	778	11.4	435	5 08S2M1	08S2M1 drosophila
27	754.5	11.1	311	13 09PV24	09PV24 xiphophorus
28	754.5	11.1	1362	13 09PV24	09PV24 xenopus lae
29	734	10.8	331	4 09BUD7	09BUD7 homo sapien
30	733	10.8	1671	5 09NJV5	09NJV5 biophalari
31	723	10.6	149	6 09B666	09B666 oryctolagus
32	705.5	10.4	1368	13 08UM85	08UM85 paratichthy
33	699	10.3	1418	13 093457	093457 scophthalmu
34	682.5	10.0	1369	13 08UM86	08UM86 paratichthy
35	679	10.0	1358	13 073798	073798 xenopus lae
36	668.5	9.8	1472	5 09US48	09US48 bombyx mori
37	658	9.7	1412	13 08UM84	08UM84 paratichthy
38	645.5	9.5	1418	13 08UM83	08UM83 paratichthy
39	645	9.5	1245	13 09YGH8	09YGH8 scophthalmu
40	641	9.4	1371	11 09QVW4	09QVW4 rattus sp.
41	623.5	9.2	2144	5 09VDP4	09VDP4 drosophila
42	597	8.8	987	11 091YMO	091YMO mus musculu
43	596	8.7	935	4 096L35	096L35 homo sapien
44	594	8.7	987	11 099MR2	099MR2 mus musculu
45	587.5	8.6	1036	4 007912	007912 homo sapien

## ALIGNMENTS

RESULT 1  
ID 018735 PRELIMINARY; PRT: 1259 AA.  
AC 018735  
DT 01-JAN-1998 (TEMBLrel. 05, Created)  
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE Brdb-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "cDNA cloning of erdb-2 from canine mammary gland."  
RT Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.  
RL EMBL; AB008451; BAA23127.1; -  
DR HSSP; P11362; IFGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR Pfam; PF00001; Euk\_pkinase; 1.  
DR ProDom; PD000001; FU; 3.  
DR SMART; SM00261; Tyrc; 1.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS00109; Tyrosine-protein kinase.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D9C4ACD46 CAC64;

ery Match 92.4%; Score 6295; DB 6; Length 1259;  
st Local Similarity 92.4%; Pred. No. 0;  
ches 1160; Conservative 39; Mismatches 55; Indels 6; Gaps 2;  
1 MELALCRWGLLALPPGAATQVCTGDMKRLRASPETHLDMRLHYOGCQVQGNL 60  
1 MELAMCRWGLLALPPGAATQVCTGDMKRLRASPETHLDMRLHYOGCQVQGNL 60  
61 ELTYLPTNASSLFLDIOEVQGYVLIANNQVRLPQRLIRVGTQLFEDNALVLDNG 120  
61 ELTYLPTNASSLFLDIOEVQGYVLIANNQVRLPQRLIRVGTQLFEDNALVLDNG 120  
121 DPLNNTTAVTASPGGLRELQRLSTELLKGVLIQRNPOLCYDTITMKDIFHNQOLA 180  
121 DPLNNTTAVTASPGGLRELQRLSTELLKGVLIQRNPOLCYDTITMKDIFHNQOLA 180  
121 DPLNNTTAVTASPGGLRELQRLSTELLKGVLIQRNPOLCYDTITMKDIFHNQOLA 180  
181 LTLIDTNRSAKCHPCSPKCSKCSKCSKCSKCSKCSKCSKCSKCSKCSKCSKCSK 240  
181 LTLIDTNRSAKCHPCSPKCSKCSKCSKCSKCSKCSKCSKCSKCSKCSKCSKCSK 240  
241 AAGCTGKSHDCLALPHNHSICELHCPALVTYNTDTFESMPNPEGRTGASCTACP 300  
241 AAGCTGKSHDCLALPHNHSICELHCPALVTYNTDTFESMPNPEGRTGASCTACP 300  
301 YNYLSTDVGSCTLVCPAHNCEVTAEDGTQRCCKSKCARVCYGLGMEHLREVAVTAN 360  
301 YNYLSTDVGSCTLVCPAHNCEVTAEDGTQRCCKSKCARVCYGLGMEHLREVAVTAN 360  
361 IOEFACCKKIFGSLAFIPESFDDPASNTAPLOEOLQVETLEITGYLYISAMPDLP 420  
361 IOEFACCKKIFGSLAFIPESFDDPASNTAPLOEOLQVETLEITGYLYISAMPDLP 420  
421 DLSVFNLOVIRIRIANGAYSLTLOGLISWLGSLRLSRLSGLALIHNTLCPVHTV 480  
421 DLSVFNLOVIRIRIANGAYSLTLOGLISWLGSLRLSRLSGLALIHNTLCPVHTV 480  
481 PWDOLFNRPHQALHTANPDECEGEGALCHQOLCARHCWGPQVCNCSQFLRGQBC 540  
481 PWDOLFNRPHQALHTANPDECEGEGALCHQOLCARHCWGPQVCNCSQFLRGQBC 540  
541 VEECEVLOGLPREYVNAHRLCPHCECQPNQSVTCFGEADOCVACAHYKDPFCVANC 600  
541 VEECEVLOGLPREYVNAHRLCPHCECQPNQSVTCFGEADOCVACAHYKDPFCVANC 600  
601 PSQVPLDLSWPIKMFPEDEGACQPCPINCSTHSCVDLDKGPAPQASPLTSIYAVNG 660  
601 PSQVPLDLSWPIKMFPEDEGACQPCPINCSTHSCVDLDKGPAPQASPLTSIYAVNG 660  
661 ILVAVLVGVVGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
661 ILVAVLVGVVGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
721 RKXKVLGSGAFVYKGIWIPDGENVKIPVAIKVRENTSPKANKELIDEXYVNAVGSP 780  
721 RKXKVLGSGAFVYKGIWIPDGENVKIPVAIKVRENTSPKANKELIDEXYVNAVGSP 780  
781 YSRLLGICLTSTVQVLTQMLPQCLDHYREHGRGLGSDLLMWCQIAKGSYLEDDR 840  
781 YSRLLGICLTSTVQVLTQMLPQCLDHYREHGRGLGSDLLMWCQIAKGSYLEDDR 840  
841 LVHRDLAANVAVKSNHAKITDGLARLLIDETETHADGKAVPIKMALESILRRFT 900  
841 LVHRDLAANVAVKSNHAKITDGLARLLIDETETHADGKAVPIKMALESILRRFT 900  
901 HOSDWSYVAVTWEMLTFGAKPYDGIIPAREIPDLLEKGERLPQPCITTDVYIMKXCM 960  
901 HOSDWSYVAVTWEMLTFGAKPYDGIIPAREIPDLLEKGERLPQPCITTDVYIMKXCM 960  
961 IDSECRPFRELIVSEFSRARDPQRFVYQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020  
961 IDSECRPFRELIVSEFSRARDPQRFVYQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020  
1021 EBYLVPOGFCPCDPAPGAGVHHRHSSSTRSGGDLTLGLFSEEBEAPRSLAPSEG 1080

Db 1020 EBYLVPOGFCPCDPAPGAGVHHRHSSSTRSGGDLTLGLFSEEBEAPRSLAPSEG 1080  
Qy 1081 AGSDVDGDLGMAAGLQSLPTHDSPLOVSEDPVPLPSESTDGVAPLTSPOREYV 1140  
Db 1080 AGSDVDGDLGMAAGLQSLPTHDSPLOVSEDPVPLPSESTDGVAPLTSPOREYV 1140  
Qy 1141 NOPDVRPOPSPREGPLPARPAGATLER-----AKTSPKXGVKDVAFAGAVENPE 1195  
Db 1140 NOPDVRPOPSPREGPLPARPAGATLER-----AKTSPKXGVKDVAFAGAVENPE 1195  
Qy 1196 YLTPGGAAPQHPHPPAPFADNLVYMDQDPPEERGAPOSTKGTPTAENPEYGLDVEV 1255  
Db 1200 YLTPGGAAPQHPHPPAPFADNLVYMDQDPPEERGAPOSTKGTPTAENPEYGLDVEV 1255  
RESULT 2  
ID 090X70 PRELIMINARY; PRT; 1209 AA.  
AC 090X70;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Epidermal growth factor receptor.  
GN EGFR.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OX Mammalia; Euteleota; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1] NCBI\_TaxID=10116;  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RX MEDLINE=90258888; Pubmed=2342466;  
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
RT "A truncated, secreted form of the epidermal growth factor receptor is  
RL encoded by an alternatively spliced transcript in normal rat tissue."  
RN [2] Mol. Cell. Biol. 10:2973-2982(1990).  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RA Petch L.A.;  
RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RA Gutridge K., Dawson T.L., Earp H.S.;  
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
RN [4] EMBL; M37394; AAF14008.1; -  
DR HSSP; P11362; IFCK.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR002174; Furin\_kinase.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR Prodom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR PROSITE; PS00219; TYK; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_DOM; 1.  
DR ATP-binding; Receptor; Tyrosine-protein kinase.  
SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B7773 CRC64;  
Query Match 46.6%; Score 3171; DB 11; Length 1209;  
Best Local Similarity 50.3%; Pred. No. 26-230;  
Matches 642; Conservative 167; Mismatches 354; Indels 114; Gaps 26;  
3 LAALCRWGLLALPPGA-ASTQVCTGDMKRLRASPETHLDMRLHYOGCQVQGNL 61

```

Db      15 LAALCAAG-----GALKEKKVCOGTSNRRLTQLTFEDHFLSLQRMNCEVVLGNLE 66
Qy      62 LTYLPTNMSLSTLODIQEVQGYVLIAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
Db      67 IYVORNYDLSFLKTIQEVAGYVLIALNTERIPLENIQIRGNALXYENTYALAVLSN-- 124
Qy     122 PLNNTPTVPGASPGGIREQLRSLTEILKGVLIQGNPOLCYODTILMDIFHKNNQLAL 181
Db     125 -----YGNKTKYRELPMRNLOEILGAVFSPNNPILCNNETIQQMDIV-QDVFLSN 175
Qy     182 TLIDITNRS-RAGHPGSPMKSGSRGMESESEDCOSLTRVCAGGCA-RCKGPLPTDCCHBO 239
Db     176 MSMDOVRHLITGCRKDCBSPGNSCKRGSENGCKLTIKIICAQOCSRCRGSRSDCCCHAO 235
Qy     240 CAAGCTGPRHSDCLACLAHFNHSGICELHCPALVTYNTDTFESMNPDEGRYTFGASCVTAC 299
Db     236 CAAGCTGPRSDCLVCHRFREDBATCKDTCPMLLYNPPTYQMDVNPDEGRYTFGASCVTAC 295
Qy     300 PNYVNSTDVSGCTIVCPRLHNOEVTAEADGROCEKSKPCARVYCYGIGMEHLEVRAYTSA 359
Db     236 PNNYVNTDHGSCVRACGPDYEV-EEGVSKCKCKGPRKVCNGIGIGEFKDTLSINAT 354
Qy     360 NIOEPAGCKKIFGSLAFLEPESFDGPASNTAPLOPEQLQVFEETLEITGYLYISAMPDSL 419
Db     355 NIKHFKYCTAIISGDHLILPVAFKGDSFTRTPLDPRELEILTKYKEITGFLIQAAMPENW 414
Qy     420 PDLSPFQNIQVIRGRILNNGAVSLTQIGISMLGRSIRELGSGLALIHNNHTLCEVAT 479
Db     415 TDLHAFENLEIRGRTKQHGFSLAVVGLNITSLGRSLKEISDGVIIISGNRLCYANT 474
Qy     460 VPMQDLFNRPHQALHTARPREDECEGELACHQICARHCNMGPGTQCVNCSQFLRGOE 539
Db     475 INMKLFGTPNOKTKIMNNRAEKDCATNHNVCNPLCSSGSGGPEPTDVCSONVSRGRE 534
Qy     540 CVEECRVLQGLREYVNAHCLPCHPECOPOGNSVTCPFPBADQACAHYDPPRCVVR 599
Db     535 CVDKNILIEGEPREVENSECICQCHPECLPOMNITCTRGDNCKKCHHYVDGPICVAT 594
Qy     600 CPSPGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDLDKCCPAPQORASP-LTSIVSAV 658
Db     595 CPSPGIMGENNTL-VMKFADANNVCHLCHANCTYGCAGPLKGC-QQPEGPKIPSIATGI 651
Qy     659 VGLILVVLGVVFGI-LIKRROOKIRKYMRRLDTELVPELTPSGAMPNOQORILKE 717
Db     652 VGGLLFIV-VALGIGLFEMRRQLVAKRRLRLDERELVEPLTPSGEAPNNAHRILE 710
Qy     718 TELRKVYLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKXKEILDEAYVMAGV 777
Db     711 TEFKKIVYLGSGAFGTVYKGLMIPGEKKKIPVAIKELREATSPKANKIILDEAYVMASV 770
Qy     778 GSPVYSRLIGICTSTVOLVTQLMPYGCLLDHVRENRLGSGODLLMCMQIAKGMSTYLE 837
Db     771 DNPVYCRLLIGICTSTVOLITQLMPYGCLLDYVREHKDNIQGYLLNMCVQIAKGNVYLE 830
Qy     838 DVRLVHRDLAANVLYKSPNHKIKITDFGLARLLDIDETVHADGKVPKIMWALSILRR 897
Db     831 DRLVHRDLAANVLYKTPQHVKITDFGLAKLGAEEKYHAEGGVPIKIMWALSILHR 890
Qy     898 RFTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPOPICTIDVYMIWK 957
Db     891 IYTHOSDVMSYGVTVWELMTFGSKPYDGIPIASEISILEKGRLLPOPICTIDVYMIWK 950
Qy     958 CMMLIDSECRPRFRELIVSEFSRMAPDQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDKD 1016
Db     951 CMMLIDADSRPKRELILFEFSKMAPDQRFVVIQOGERMMLPAPTDSNFYRALMEEDMD 1010
Qy    1017 LVDAEYVLPQGGFCPRDAPAGAGMHHHRSSSTRSGGDLTLCLEFSEEARPSPLA 1076
Db    1011 VVDADYEYLIPQGGF-----NSPST-----SRTPLL 1036
Qy    1077 PSEGASGVDFDGLCMGAAGKGLSLPTHDPSPLQRYSEDPVLPSPET--DGVVAPLITS 1134
Db    1037 SSLGANSN-----SSIVACINRGSGCVKEDATFLQYSSDPTSVLTIEDNIDTFL----- 1086

```

```

Qy     1135 POPEYVNPDPVRQPPSPREGPLPAARPGATLERAKTISPGKGVYKQVFAFGAVENP 1194
Db     1087 PVPEYINQ-SVPRKRPAGSVQNPYTHNQPLHP-----APGRDLHYQN--PSHNAVSNP 1135
Qy     1195 EYL-TPQGAAPQHPHPSPAPFNDLYYWDQ-----DP-----PENGAPSTF 1237
Db     1136 EYLTNTAQ-----PTCLSSGFDSSALMWIOGSHQMSLDNDYOODFFPKAKPWGIF 1186
Qy     1238 KGTPEANPEYLGIDVP 1254
Db     1187 KG-PTNENAEYLRVAP 1202

```

## RESULT 3

```

Q9EP98 PRELIMINARY; PRT; 1210 AA.
ID Q9EP98
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Scheel C., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

```

```

DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG28045.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; Cytochrome b5.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR01245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```



SEQUENCE 1165 AA; 129614 MW; 7F7E38D8771A7AE CRC64;  
Query Match 40.4%; Score 2751; DB 13; Length 1165;  
Best Local Similarity 45.9%; Pred. No. 1e-198;  
Matches 584; Conservative 163; Mismatches 385; Indels 140; Gaps 29;  
QY 1 MELALACRWGLLALLPFG-AAST-----OVCSTDMKRLRLPASPETHLMLRHLYOGCOV 55  
DB 4 LELLEL-----LILLLLLSIGRCSTDPDRKVCOSTNQMTM-----LDNHVLYKMKMYSGCNV 56  
QY 56 VQGNLELYLPFNASLFLQDIOEVGYVLIANNOVROVLPORLRIVRTOLEFEDNYALA 115  
DB 57 VLENLEITYTOENQDLSFQSIQEVGYVLIANNEVSTIPLVNLRLIRQONLYEGFTLL 116  
QY 116 VLDNDPLNNTTPYVGASFGRLRLSLTEILKGVLIQRNPOLCYODTILMKDIFPK 175  
DB 117 VMSNVOK-NPSSP--DIVQVGLKQQLSLTLTEILSGVAVSHNPILCNVETIMMDIYVK 173  
QY 176 NNQALTLITNRSRACPCSPCKSGRCWGESSEDDQSLTRTVACAGG-ARCKGPLPTD 234  
DB 174 TSNPTMNLIPHAFEROCCKCDPGCVNGSCWAPGPHCQKFTKLCAEQCNRCRCRGPID 233  
QY 235 CCECCAGCTGPKKSDCLACHPNHSICELHCPALVTYNTDPESMNPREGRTYFGS 294  
DB 234 CCEHCAGCTGPRATDCLACRDFNDGTCKDTCPPKIYDIVSHQVNDNPNIKTFGA 293  
QY 295 CTTACPYNYLSTDVSCITVCPLEHNOEVAEDGTORCEKSCPKCARCYGLCMEMHLEVR 354  
DB 294 CYKECPSNVYVTE-GACVRSACAGMLEVD-ENGKSCCKCDVCPKVCIGIGISLSTNI 351  
QY 355 AVTSANIOEPACCKKIFGSLAFLPESFGDDPASNTAPLOPQOLQVETLEETITGYLYSA 414  
DB 352 AVNSTINISFSNCTINGDIILNRNSFEGDPPHYKIGPMPDEHMLNLTVEKITGYLVIMW 411  
QY 415 WPDLSLPDLSVPONLOVIRGRILHNGAYS-LTIQGLISWLGRLSRLSGLALHNMH 473  
DB 412 MPENNTSLSVFQNLLEIRGRITFSRGSFVVVQVSHLQWLGSLSLKVEVAGVILLKNTPO 471  
QY 474 LCFVATVPMDDOLFRNPHQALHTANRPEDECVGEGLAQHLCARGCMGPRFTQCVCNSO 533  
DB 472 LRYASTIMRRLFRFEDOSIEVDART-----ENQTCNNESEDDCWMGPGTMCVSCIAH 524  
QY 534 FLRGQECYEECRVLOGLPREYVNAHRCLEPCHEPCOPONGSVTCFGEADQVACAHHYKP 593  
DB 525 VDRGRCVACSNLLGGEPRFAQVDRGVQCHQECIVQTSITCYGGRGANCKCAHPQGG 584  
QY 594 PFCVARCDSGVKPDLSYMPIMKFPDEBGCORCPINCHSCVYDLDDKCPAQRASPLTS 653  
DB 585 PQCIPTCPHGMIGDGDTL-IMKYADKMGQCCQPCCHQNCCTQGCSPGLSGCRGD-IVSHSL 642  
QY 654 IVSAVVGILLVVLGVFEGLIKRRQOKIRKTYMRLLOETELVEPLTPSGAMPNOAQR 713  
DB 643 AVGLVSGILLITYVALLIVLIRRRIRK-RKRTIRLLOEKRLVERBLTPSGAPNOAFLR 701  
QY 714 ILKETLRKVKVLSGAFGTVYKGIWIDGENVKIPVAILKVLRENTSPRANKELIDEAV 773  
DB 702 ILKETEFKKDRVLSGAFGTVYKGLMNPGENIRIPVAILKVLRENTSPRANKELIDEAV 761  
QY 774 MAGVSPVYSRLGICLTSTVQVLTQMLPYGCLLDHVENRRLSODLLMCMQIAKGM 833  
DB 762 MASVDHPVRCRLGICLTSAVOLVQMLPEYGLLDYVRHORIRIGQWLLMNCVOIAKGM 821  
QY 834 SYLEVDRLVHRDLAARNVLSVPHVKITDIFGLARLLIDETEVHADGGKVPKIMMALES 893  
DB 822 NYLEBRHLVHRDLAARNVLSVPHVKITDIFGLARLLIDETEVHADGGKVPKIMMALES 881  
QY 894 ILRRRFTHQSDVSYGVTVMEMLATFGAKPYDGIIPAREIPDLLEKEBRLPQPICTIDVYM 953  
DB 882 ILQWYTTQSDVSYGVTVMEMLATFGSKPYDGIIPAKETIASVLENERBLPQPICTIBVYM 941  
QY 954 IMVKCMMDISECRPFRELYSEFSRMARDPQRFVVIQWEDLGPAPLPDSTFRSLLEDD 1013  
DB 942 IILKCMMDIPSSRFRELYGEFSQARDPSRYLVIQG--NLPSPSDRRLFSRLSSDD 998

QY 1014 MGDLYDAEELVYPOQGFPCPDPAAGAGVWHHRSSSTRSGGDLTLGLEPSEEBAPRS 1073  
DB 999 --DIVDADAEYLL-----RYKRIN-RQSS-----E 1019  
QY 1074 PLAPSEGAGSDVFDGDLGMAKGLQSLPTHPSPLOKYSBDPTV-PLPSETDGYAPLT 1132  
DB 1020 PCIPRNGH-----PVRENSILRYSIDPTQNALEDLDLGH----- 1054  
QY 1133 CSPQPEYVNPDPVROP-----PSRE-----GPLP-AARPAGTLERAKTSLSPCKNG 1179  
DB 1055 ----EYVNPQSESTSSRLSDIYNPYEDLTQMGVSVLSQEAETFSRPEVLTNNQNS 1109  
QY 1180 VVKDVFAFGAVENPEYLTPOGGAAPQHPAPPAFPNLVYWDODPPERGAPSTFKG 1239  
DB 1110 L--PLVSSGSDDDPY---QAG-----YQAAF-----LQQTGALTONGMF 1144  
QY 1240 TPTAENPEYGL 1251  
DB 1145 LPAENLELYGL 1156  
RESULT 5  
Q9W6F6 PRELIMINARY; PRT; 1137 AA.  
AC Q9W6F6;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Receptor tyrosine kinase (Fragment).  
GN ERBB4.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
NCBI\_TaxID=9031;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=HINDRAIN;  
RX MEDLINE=99263203; PubMed=10328884;  
RA Dixon M., Lumsden A.;  
RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in  
RT embryonic chick hindbrain."  
RL Mol. Cell. Neurosci. 13:237-258 (1999).  
DR EMBL: AF121963; AADJ1764.1; --  
DR HSSP: P11352; 1FGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_Pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001368; TNFR\_C6.  
DR InterPro: IPR001245; Tyr\_Pkinase.  
DR InterPro: IPR004019; YLP\_motif.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 1.  
DR Pfam: PF02757; YLP; 2.  
DR PRINTS: PR00109; TYRKINASE.  
DR ProDom: PD000001; Euk\_Pkinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; Tyrc; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS00652; TNFR\_NGFR\_1; UNKNOWNV\_1.  
KW Kinase; Tyrosine-protein kinase.  
FT NON\_TER  
SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC64F CRC64;  
Query Match 40.1%; Score 2729.5; DB 13; Length 1137;  
Best Local Similarity 47.3%; Pred. No. 4.1e-197;  
Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;  
QY 161 LCYQPTILMKDIFHNQNLALTLITNRSRACPCSPCKSGRCWGESSEDDQSLTRTVAC 220





```

QY 421 DLSVQNLGVNIGRIILHNGAYSLTLQGLGISTWLSRLSELGLALIHNNTHLCFVHTV 480
DB 404 DLSVSSLTTLTGSRSLFKRSLMVMRIPLTSLGLSLREISGSVYISQNALCYHTTV 463
QY 481 PMDOLFRNPH-QALLHTARPEDECVGELAGHQLCARHGCWGPRTQCVNCGQFLRGE 539
DB 464 NMTQLFRSGRVANSLSNRPMABCVADRVCPLSDSGCWGPGPDQCLSCNNYSRHGT 523
QY 540 CVEGCRVLQGLPREYVNAH-CLPCHPECOQNGSVTCFGEADQCVACAHYKPPFCVA 598
DB 524 CVAGCHFNGIIPREFAGLNGVACVACHBECKPQTGKASCTGPGADBECACTKPRDGYCMS 583
QY 559 RCPGSKPDLSTYMPKPFDEGACOPCPINCTHSCVDLDDKCCPAEQRASPLTISAV 658
DB 584 SCPAGVN-DGEXGLIFKPFNRGHCERCHQNCQSGCGPLNCC--LEAARLTISGQI 639
QY 659 VGLLVVVLGVF-----GLIKRQOKIRKYTRRLQETELVEPLTPSAMNQAO 711
DB 640 TGIAGVAGLIFCLVLPFLGMLYHGLAIRRQAKRRYLESSESEPLGP-GEKGTKVH 698
QY 712 MRLKETELRKVKVLGSGAGTVYKGIWIPDGENVKIPYAIKVLRENTSPKANKELIDBA 771
DB 699 ARLKPSDLRKIKPLSGGVFGTVSKGFWIPBETVXIPVAKITIDSSGRQTFTEITDL 758
QY 772 YMAVGSPYVSRLLIGITLSTVQLVTQMLPYGCLLDHVENKRLGSDLLMCMQIAK 831
DB 759 LSMGSLDHPYIVRLIGICPGTCLQIATQLSHSGSLLEHTRQHTSIDPRLINMCVQIAK 818
QY 832 GMSYLEDVRLHARDLAARVLVKSPMHVKITDGLARLIDIDETEHAOGKRPITMML 891
DB 819 GMYVEEHVVKHNLAKRNLKNDYQVQISDGVADLLYPDKKYVSEYETKPIKMML 878
QY 892 ESILRRRFTQSDVNSYGVTVMLMFGAKPYDGIIPAREIPDLLEKGERLPOPICTIV 951
DB 879 ESILRRRFTQSDVNSYGVTVMLMFGAKPYDGIIPAREIPDLLEKGERLPOPICTIV 938
QY 952 YMIWKCWMIDSECPREFELVSESRMARDPQRFVIVNEDLGPASPLDSTFFYRSLID 1011
DB 939 YMWVKCWMIDENIRPTFKELASDFTRMARDPRYLIVMEG-----ED 982
QY 1012 DDMGDLVDAEVLVQGGFFCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEAR 1071
DB 983 SGMGFFL-----RQSER--GLLEADLEEDDEE-- 1008
QY 1072 RSLPAPSEGAGSDVFDGLGMC--AAKGLQSLPTHDPSPLO-----RYSEDPVY 1118
DB 1009 -----GLDRAFTPSLOPSPSMSTSPSQINSYMTQLRYD----- 1044
QY 1119 PLPSETDGVADLTCSPOP-EYVNO-----PDVROPSPSREGPL--PAAR 1161
DB 1045 -FAVSGGHIQYLPMSPSFVDITROLWYQSRSLSVRTLPRDSAFRRSSREAELEDGAQ 1103
QY 1162 PAGATLERAKTSPKNGVAVKVFAGCAVENPEYITPQCGAAPQHPHPPASPAFDNLY 1221
DB 1104 CAGIRVR-----FGSERGN-----PQGG----- 1122
QY 1222 YMDOPPERGAPSTFGTPTAENPE 1247
DB 1123 --QQRKSTASSPSFKTMADEBDE 1146

```

RESULT 7

```

Q9BIH9 PRELIMINARY: PRT: 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

```

```

OC Anopheles.
OX NCBI_Taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ301655; CAC35008.1; -.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000345; Cyt_c_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00757; Furin-1like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR KW Receptor.
FT NON TER
SQ SEQUENCE 1433 AA; 159585 MM; E3D9D88967724F07 CRC64;

Query Match 30.0%; Score 2044.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred. No. 3; 2e-145;
Matches 473; Conservative 196; Mismatches 393; Indels 385; Gaps 36;

QY 26 CIGTDKRLPASPETHMLHLNLYGCGVQGNLETLPTNASLSFLDIOEGVGYL 85
DB 1 CIGTKRMSPANREHYKQLDRYNTCYVDGNLSTIYQITDNLFIQHIREVGYL 60
QY 86 IAHNVQVPLQRLRIVRGTOLE--EDNYALAVLDNGDPLNNTTPYVAGSPGRL 140
DB 61 ISLYDLPQVILPRLQIIRGRITFKLNMKEAVGLFV-----SFSMNTL 104
QY 141 QRLSTELLKGYLQIORNQOLCYQDTILMKDI-FHKNQDLATLIDTNSRACHPCSPMC 199
DB 105 ELPALRDIIGSGVGFENNYNLCHKMSINWEEILAPQTSMQYTFNFPSSPERVCPCHPSC 164
QY 200 KGSRCWGESSEDCQSLTRIVCAGGCA--RCKGPLPTDCCHEOCNAACCTGPKHSDCLACLH 257
DB 165 EVG-CMGBGANHCQRPSTKNTSPQSGRCFGKPRECHLFCAGGCTGPTGSDCLACKN 223
QY 258 FHSIGICELHCPALVTYNTDFESMPNREGRYTFGASCVYACPYNTLSTDVGSCTLVCP 317
DB 224 FYDDGVCKQECRPMQYNTNPTNYFWEBNPDOKAYATGYRCKP-ELLLDNGACVAKCPK 282
QY 318 HNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREYRAVTSANTIOFPACKKIFGSLAFL 377
DB 283 GMPQONS-----CVPCKGVCPKTCPEGIVH-----SDNIGYKQCTIIEGSLLEI 329
QY 378 PESFGDPAASNT-----APLOPQLOVFEETLEITGVLVYSAMPDSLPLDSVQNLQ 429
DB 330 DQSFQFOQVYNTNFGPRYIKIDPRLVFTVKEITGFINIOAHNPFTTLNRYRNE 389
QY 430 VIRGILHNGAY-SLTQGLGISWLSRLSELGLALIHNNTHLCFVHTVPMDOLEFRN 488
DB 390 VVGGRQKLENLPLASVYIVTSLKSLLEKSKLVNNGSIYIILNSDCLFEDIDMSIKKS 449
QY 488 PQOALLHTARPEDECVGELAGHQLCARHGCWGPRTQCVNCGQFLRQGEVCEGRVLO 548
DB 450 SDHEVMVQKNRNVATECHGEWMECSBQSGWAGCMGKRPQGLCKKIVKYGKCLDSCK--- 506

```

```

OY 549 GLPREY-VNARHCLPCEPCOPONGSVTCFGEADQCVAAHYKDPFVCARCP----- 601
DB 507 SLPLRYSDSKTCGCHOCCKD-----FCYGNEDNCGSCMNVKGRFVVAECPTTKAM 561
OY 602 -----SGVPRDLSYMPIMKPPD----- 618
DB 562 NGCTCINCHTKVCGRGRPTDTIAPDGCISCDKAIISDAKIERLMDESCPDYSDYVL 621
OY 619 -BEG----- 621
DB 622 QEBGRLKOLSGKAVCKRCHPRCKCTGYGFHFOECOECTGYKKGEDECPDQFYANEE 661
OY 622 --ACOPCPINCT-----HSCVDL-----DD-----KCGPAEQ----- 646
DB 682 TRCLPCHOCRCGCHGAGDHDHCCRNKLREGDPDATTFTCVSNCPASHPRKPRQEA 741
OY 647 -----RASPLTSIVSAVGLLVVVLGVVFGI---LIKROOKIRKXTM 687
DB 742 GKIGPYCSADSMOSGLIEPOTQVKIYMGVSMALJLCCVFGIAFVLFNRKKNKDAVAM 801
OY 688 RRLLOETELVEPLTPSGAMPNOAQMRILKETELRKVKVLSGAFGYKGIWIDGENVK 747
DB 802 TMAIAGEDESEPLRPSVNGPLTLRIIKEAIRRGVLCMGAFGRVFKGVMBEGESVK 861
OY 748 IPVAIKVLRENTSBRKANKELIDENYVMAVGSFYVSRLLGICTSTVQLVTOLMPYGLL 807
DB 862 IPVAIKVLEWGSSESSKEFLBAYIMASVHPMLIKLAAVCMTHOMMLITOLMPLGLL 921
OY 808 DHVRENGRLSODLLMCMQIAGKMSYLEVRLVHRDLAARVLYKSPNHVKTIDFGILA 867
DB 922 DYVRNNKOKISKALNNSTOJARMATLEERLVRDLAARVLYQPSCVKIKIVFGILA 981
OY 868 RLLOIDETEHADGKVPKIMMALESILRRRFTHOSDVASGVYVWELMTFSAKYYDGI 927
DB 982 KLBDSDSEYRAAGAKMIRIMALECIRHVRFTSKSDVAFGIITWELTYGARPEVNP 1041
OY 928 AREPLDLKEGRLPQRPICITIDVYIMVWKCMIMSECPREPRELVESESRMARPOREV 987
DB 1042 AKDVDELIEIGHKLPQPCISLDVYCILLSCVLDADARPTKOLAETABAKARDGRYL 1101
OY 968 VIQNEDELASPLDSFFYRSILLEDPMGLV----- 1018
DB 1102 MI-----PODKRMRLPSTTNODEKOLIRTLAVMAAAAAAGASNVDPSTIA 1152
OY 1019 DAEEYLVPOQGFPCPPAPAGAGVHRRHSSSTRSGGDLTLGLEPSEBEARS----- 1073
DB 1153 EIDVEYLOPKTRPSIMLGPSSA-----VEPS--DEMPKSLRYCK 1188
OY 1074 -PLAP--SEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVTLPSETDGYVA 1129
DB 1189 DPLKPDDETDGHGKEY-----GVGGIR-----IMLPDDEDDYLM 1222
OY 1130 PLTCSQREPVVNGPDRPQPPSPREGPLPARPAGATLERAKTSLPGKNGVNDVFAFG 1189
DB 1223 P-TCOSO-----NOS-----TFG-----YMDLIGVPA 1243
OY 1190 AVENPEYL-----TPQGAAPQPHPPAPSPAFDNLTYWDQDEPPEKAPSTFKGT 1240
DB 1244 SVDNPEYLMSTQAIAGLAQOSMG--PHTRP-----PNTENGW 1280
OY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

```

RESULT 8
O9UK79 PRELIMINARY; PRT; 419 AA.
AC O9UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstatin.

```

```

GN HER-2.
OC Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=9415951; PubMed=10485918;
RX Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RA "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF177761; AAD56009.2;
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L domain; 1.
DR SMART: SM00261; Fu; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECLIBE347E2D030C CRC64;

```

```

Query Match 27.5%; Score 1874.5; DB 4; Length 419;
Best Local Similarity 89.6%; Pred. No. 3,66-133;
Matches 353; Conservative 5; Mismatches 29; Indels 7; Gaps 2;

```

```

OY 1 MELALCRWGLLALLPFGAASVCTGTDKRLIPASPEHLDMLRLHYGGCVVQGNL 60
DB 1 MELALCRWGLLALLPFGAASVCTGTDKRLIPASPEHLDMLRLHYGGCVVQGNL 60
OY 61 ELTYLPTNASLFLDIOEVQGYVLIANOVROVPLORLIVRGQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLFLDIOEVQGYVLIANOVROVPLORLIVRGQLFEDNYALAVLNG 120
OY 121 DPLNNTTPTVTCASPGGLRELRLSLTEILKGGVLIQRPOLCYOTIILMKDI FHKNOQA 180
DB 121 DPLNNTTPTVTCASPGGLRELRLSLTEILKGGVLIQRPOLCYOTIILMKDI FHKNOQA 180
OY 121 DPLNNTTPTVTCASPGGLRELRLSLTEILKGGVLIQRPOLCYOTIILMKDI FHKNOQA 180
DB 121 DPLNNTTPTVTCASPGGLRELRLSLTEILKGGVLIQRPOLCYOTIILMKDI FHKNOQA 180
OY 181 LFLIDTNRBRACHPCSPMKCKSGRCESEDCQSLTRYVACGACRCKRPLPTDCCHEOC 240
DB 181 LFLIDTNRBRACHPCSPMKCKSGRCESEDCQSLTRYVACGACRCKRPLPTDCCHEOC 240
OY 241 AAGCTGPKXSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCTYAC 300
DB 241 AAGCTGPKXSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCTYAC 300
OY 301 YNTLSTDVSGCTIVCGELHNOEYTAEDGTORCEKSKRCARVVCYGLGMEHLREYAVTSAN 360
DB 301 YNTLSTDVSGCTIVCGELHNOEYTAEDGTORCEKSKRCARVVCYGLGMEHLREYAVTSAN 360
OY 361 IOEFAGCKKIFGSLAFLESFDDDPASNTAPLOP 394
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSAPSLPAP 387

```

```

RESULT 9
O8R2X1 PRELIMINARY; PRT; 367 AA.
ID O8R2X1;
AC O8R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP Straube R.;
RA Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.

```

DR EMBL/BC027080.1; AAH27080.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;  
 Best Local Similarity 88.0%; Pred. No. 5.1e-123;  
 Matches 333; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPAREIPDLEKEERLPQPPIC 948  
 DB 1 MALESILRRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPAREIPDLEKEERLPQPPIC 60  
 QY 949 IDVYIMVCMWIMDSECRPFRELVSFSGMADPQRFVITQMEDLGASPLDSTYRSL 1008  
 DB 61 IDVYIMVCMWIMDSECRPFRELVSFSGMADPQRFVITQMEDLGASPLDSTYRSL 120  
 QY 1009 LEDDDMGDLVDAAEYLVPOGFCPPAPAGAGMHHRRSSSTRSGGGLTLGLPSEE 1068  
 DB 121 LEDDDMGDLVDAAEYLVPOGFCPPAPAGAGMHHRRSSSTRSGGGLTLGLPSEE 180  
 QY 1069 EAPRSPFAPSEGAAGSDVFDGLGMAKGLQSLPTHDPSPLOYSSEDPVLPFSETDGYV 1128  
 DB 181 EAPRSPFAPSEGAAGSDVFDGLGMAKGLQSLPTHDPSPLOYSSEDPVLPFSETDGYV 240  
 QY 1129 APLTCSPOREYVNPQVPRQPPSPRSGPLPAAPPAATLERATLSPGKGVKDVAFAG 1188  
 DB 241 APLTCSPOREYVNPQVPRQPPSPRSGPLPAAPPAATLERATLSPGKGVKDVAFAG 300  
 QY 1189 GAVENEYLTPOGAPAPPPAPFAPFNLYYMODDPERGAPSTFEGTPTANPEY 1248  
 DB 301 GAVENEYLTPOGAPAPPPAPFAPFNLYYMODDPERGAPSTFEGTPTANPEY 360  
 QY 1249 LGLDVFPV 1255  
 DB 361 LGLDVFPV 367

RESULT 10  
 Q86712 PRELIMINARY; PRT; 729 AA.  
 AC 086712;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Polypeptide.  
 GN POLYPEPTIDE.  
 OS Avian rous-associated virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
 NC NCB1\_Taxid=11950;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94203659; PubMed=8152791;  
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities.";  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL; S69372; AAC60725.1; -  
 DR HSP; P03332; IAS6.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR004028; Retro\_M.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF02813; Retro\_M; 1.  
 DR ProDom; PDD00001; Euk\_pkinase; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 729 AA; 80649 MW; 8402F614FEFID63 CRC64;

Query Match 25.2%; Score 1720; DB 15; Length 729;  
 Best Local Similarity 54.8%; Pred. No. 3.9e-121;  
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 569 PQNGSTTCRPEPDQCYACAHYDPPFCVACRCSGVKVPDLSYPMKMFPEBEGACQPCPI 628  
 DB 141 PEETAPPKTGP--DHCKCAHFIDGPRCYVACAGVAGENDTL-VWYCAANAVCOLCHP 197  
 QY 629 NCTHSCVDDLDKCGCPAQASPLTSTIVSAV--GILVVTVGVFGILIKRQOKIRRYTM 687  
 DB 198 NCTHSCVDDLDKCGCPAQASPLTSTIVSAV--GILVVTVGVFGILIKRQOKIRRYTM 253  
 QY 688 RLLQETELVEPLTPSGAMPNQAOMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVK 747  
 DB 254 RLLQETELVEPLTPSGAMPNQAOMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVK 313  
 QY 748 IPAIVATVRENTSPKANKELIDEAAYMAGVGSYVSRLLGICTSTVQLVTQMPYCLL 807  
 DB 314 IPAIVATVRENTSPKANKELIDEAAYMAGVGSYVSRLLGICTSTVQLVTQMPYCLL 373  
 QY 808 DHVRENRRGLSGDILNMCMQIAKMSYLEVRLVHRDLAARVTVKSPMHVKTDFGLA 867  
 DB 374 DHVRENRRGLSGDILNMCMQIAKMSYLEVRLVHRDLAARVTVKSPMHVKTDFGLA 433  
 QY 868 RLIDIDETEHADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 927  
 DB 434 RLIDIDETEHADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 493  
 QY 928 AREIPLLEKEERLPQPPICITDVYIMVCMWIMDSECRPFRELVSFSGMADPQRFV 987  
 DB 494 AREIPLLEKEERLPQPPICITDVYIMVCMWIMDSECRPFRELVSFSGMADPQRFV 553  
 QY 988 VIO-NEDLGAPSPDSTFYSLLLEDMDGDLVNAEYLVPOGFCPPCDPAPAGAGMHHR 1046  
 DB 554 VIO-NEDLGAPSPDSTFYSLLLEDMDGDLVNAEYLVPOGFCPPCDPAPAGAGMHHR 598  
 QY 1047 HRSSSTRSGGGLTLGLPSEEAPRSP-----APSEGAAGSDVFDGLGMAKGLQSL 1101  
 DB 599 -NSPT-----SKTPLLSLSATSNNSATNCID-----RNGQH 631  
 QY 1102 PTHDPSPLORYSEDPVLPSET--DGVAAPLTCSPOREYVNPQVPRQPPSPRSGPLA 1159  
 DB 632 PTHDPSPLORYSEDPVLPSET--DGVAAPLTCSPOREYVNPQVPRQPPSPRSGPLA 675  
 QY 1160 ARPAGATLERATLSPGKGVKDV-----ARCGAVENEYLT 1197  
 DB 676 -----TANVQNIYNNISLTAISKLPMSRYNSHSTAVDPEYL 715

RESULT 11  
 Q86714 PRELIMINARY; PRT; 567 AA.  
 AC 086714;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE V-erbB protein (Fragment).  
 GN V-ERBB.  
 OS Avian rous-associated virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
 NC NCB1\_Taxid=11950;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94203659; PubMed=8152791;  
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities.";  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL; S69372; AAC60727.1; -  
 DR HSP; P11362; IFGK.  
 DR InterPro; IPR000719; Euk\_pkinase.

DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00069; kinase; 1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SM00219; TyrcK; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Tyrosine-protein kinase.  
 FT NON TER 1  
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF25E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;  
 Best Local Similarity 55.4%; Pred. No. 3.8e-121;  
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQCACAHYKQPPFCVARNCPGKVDLSYMPDIWKPRDEGACOPCPICHTSCVDL 637  
 DB 1 GP--DHCKKCAFIIDGPHCVAKCPAGVAGENDTL--VMKADANAVCOLCHPCCTGCKCP 57  
 QY 638 DDKGCPAEQASPLTSIVSAV--GILLVVLGVVFGILIKRROKIKRYTMRLLQETEL 696  
 DB 58 GLEGCP---NSKSTSIAGVVGILLCLVVGGLGLYLR--HIVKRTTLARLLQERL 113  
 QY 697 VEPLTPSGAMPNOAQRILIKETELRKVKYLSGAGFTYKGIWIPDGENVKI PVAIKVL 756  
 DB 114 VEPLTPSGEAPNOAHLRIKETEFKKVKYLSGAGFTYKGLMIPEGEKVI PVAIKEL 173  
 QY 757 ENTSKAKELIDELYVMAVGSFVYSRLGICLTSTVQLTQMLPYGGLLHVENRGR 816  
 DB 174 EATSKRANEILIDELYVMAVSDNPRVCRLLGICLTSTVQLTQMLPYGGLLHVENRGR 233  
 QY 817 LGSQDLNMCQIAKGMSTLEDVRLVHRDLAARNLVKSPNHVKITDFGLRLDIDETE 876  
 DB 234 IGSQYLLMWCQIAAGMNVLEERLVHRDLAARNLVKTPQHVKITDFGLRLDIDETE 293  
 QY 877 YHADGKPIKMALESILRRFTHOSDWSYGVTVWELMTGAKPYDGI PAREIPDLLE 936  
 DB 294 YHAEQKPIKMALESILRRFTHOSDWSYGVTVWELMTGSKYDIPASEISVYLE 353  
 QY 937 KGERLPPICTIDVYMWKCMIDSECRPFRELVSFRRMARDPQRFVYIO--NEDG 995  
 DB 354 KGERLPPICTIDVYMWKCMIDADSRPFRELIAFSGMARDPPIYVYIOGDERNA 413  
 QY 996 PASPLDSTFYRSLDEDDMDLVDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSG 1055  
 DB 414 LPSPLDSKRYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454  
 QY 1056 GGDITLGLERSEEARSPPL-----APSEGASDVVDGDLGMAKGLQSLTTHDPSPLQ 1110  
 DB 455 -----SRPLSLSLSTSNNSATNCID-----RNGGHPVREDSEFVQ 491  
 QY 1111 RYSEDPVPLPSET--DGVAAPLTCGPOPEYVNOQDVRPOPSPREGPLPAARPAQATLE 1168  
 DB 492 RYSSDPTGNFLEESIDDDGL-----PAREYVNO--LMPKRS----- 526  
 QY 1169 RAKTLSPGNQGVVKOVF-----AFGGAIVENPEYL 1197  
 DB 527 ----TAMVQNOIYNNISLTAISKLPMSDRYQNSHSAVDNPEYL 566

RESULT 12  
 Q8WYVO PRELIMINARY; PRT; 412 AA.  
 AC Q8WYVO;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 44.7 kDa protein.  
 GN PP3659.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OK NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
 RA Man D.F., Gu J.R.;  
 RT "Novel human cDNA clones with function of inhibiting cancer cell  
 RT growth.";  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF318349; AAL55856.1;  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; kinase; 1.  
 DR Pfam: PF02157; YLP; 2.  
 DR ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SM00219; TyrcK; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;  
 Best Local Similarity 80.5%; Pred. No. 8.3e-120;  
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALSEILRRFTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPICT 948  
 DB 1 MALSEILRRFTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPICT 60  
 QY 949 IDVYIMWKCMIDSECRPFRELVSFRRMARDPQRFVYIONEDLGASPLDSTFYRSL 1008  
 DB 61 IDVYIMWKCMIDSECRPFRELVSFRRMARDPQRFVYIONEDLGASPLDSTFYRSL 120  
 QY 1009 LEDDMGDLVDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEE 1068  
 DB 121 LEDDMGDLVDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEE 180  
 QY 1069 EAPRSPPLASEGASDVVDGDLGMAKGLQSLTTHDPSPLQRYSEDPVPLPSETDGYV 1128  
 DB 181 EAPRSPPLASEGASDVVDGDLGMAKGLQSLTTHDPSPLQRYSEDPVPLPSETDGYV 240  
 QY 1129 APLTSPPEYVNOQDVRPOPSPREGPLPAARPAQATLERKTLSPGNQGVVKOVFAG 1188  
 DB 241 APLTSPPEYVNOQDVRPOPSPREGPLPAARPAQATLERKTLSPGNQGVVKOVFAG 300  
 QY 1189 GAVENPEYLPQGAAPQ-----HPPPA---FSPAFLNL 1220  
 DB 301 GAVENPEYLPQGAALSTPLLLPSAQSPTSITGTRTHQSGGLHPAPSKHLRORTOST 360  
 QY 1221 YYMD--QDPPER-----GAPSTFKGTPTAEN 1245  
 DB 361 WWTQCGCEEGVRRSPDVSSGSRGLTSAGIKRMGEPPTTSRGCTGHAN 410

RESULT 13  
 Q64895  
 ID Q64895 PRELIMINARY; PRT; 962 AA.  
 AC Q64895;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Gag, V-erb-A, V-erb-B protein.  
 GN GAG, V-ERB-A, V-ERB-B.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 OC NCBI\_TaxID=11611;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9020603; PubMed=1969616;  
 RA Brunkin A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;  
 RT "Six amino acids from the retroviral gene gag greatly enhance the  
 RT transforming potential of the oncogene v-erb-B.";

```

RL  Oncogene 5:15-24(1990).
CC  -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR  EMBL; X52209; CAA36459.1; -.
DR  EMBL; X52211; CAA36459.1; JOINED.
DR  HSSP; P10828; 2N1L.
DR  InterPro; IPR000719; Euk_pkinase.
DR  InterPro; IPR000536; Hormone_rec_119.
DR  InterPro; IPR001723; Steroid_receptor.
DR  InterPro; IPR001245; Tyr_kinase.
DR  InterPro; IPR001628; Znf_Cateteroid.
DR  Pfam; PF00104; hormone_rec; 1.
DR  Pfam; PF00069; pkinase; 1.
DR  Pfam; PF00105; zf-C4; 1.
DR  PRINTS; PR00398; STRDOMONER.
DR  PRINTS; PR00047; STROIDFINGER.
DR  ProDom; PD000001; Euk_pkinase; 1.
DR  ProDom; PD000005; Znf_Cateteroid; 1.
DR  SMART; SM00430; HOL1; 1.
DR  SMART; SM00219; Tyrc; 1.
DR  SMART; SM00399; Znf_C4; 1.
DR  PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR  PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR  PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR  PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR  ATP-binding; DNA-binding; Nuclear protein; Receptor;
DR  Transcription regulation; Tyrosine-protein kinase;
DR  Zinc-finger.
SQ  SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B55CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 6,2e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEECVLTQGLPRE-VYNAR-HCLP-----CHPEQ 568
DB 354 IECQESYLAHEHYTNKKNIPHWSLTKLVADLRMIGAYHARFLMKVCECTELS 413
QY 569 PONGSVTCGPBACDVCACAHYKDPFCVACRCPGVPDLISYPIKFPDEBACQPCPI 628
DB 414 PGE-----VGP--DHCMKAHFIDGPHCVKACAPAGVLGENDTL-VKYNADANAQCQLCP 465
QY 629 NCTHSCVDLDDKGCRAEORASPLTSIVSAV-V-GILVVVLGVVFGILIKRQOKIRKRYTM 687
DB 466 NCTRCKGPGLEGCP---NGSKTPSIAGVVGGLCLVVGIGIGLYLRR-HIVKRTL 521
QY 688 RRLLOETELVEPLTSPGAMPNOAMRILKETELRKVYLGSGAFGVVYKGINIPDGENK 747
DB 522 RRLLOERELVEPLTSPGAPNOAHLRIKETEPKVKVLFQAFGVYKGLMIPDEKXT 581
QY 748 IPVAIKVLRENTSPPANKKEILDEAYVAVGVSPPVSRLLGICLTSTVOLTLQMPYGLL 807
DB 582 IPVAIKELREATSPANKKEILDEAYVAVSVDPNPHVCRLLGICLTSTVOLTLQMPYGLL 641
QY 808 DHVNRNGRLSGQDILLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHYKITDFGLA 867
DB 642 DYIREHKNIGQVYLLMVCQIAKGMNYLEERHMYHRDLAARNVLKTPQHVKITDFGLA 701
QY 868 RLIDIDETEVADGKVPKMMALLESILRRFTHOSDWSYGVYVWELMTFGAKPRDGP 927
DB 702 KQLGDEKEVHAEGGVPIKMMALLESILHRIYTHOSDWSYGVYVWELMTFGSKPDGIP 761
QY 928 AREIPDLLEKGERLPOPICTIDVYIMVYKMMIDSECRPRELVEFSFMAADPORFV 987
DB 762 ASEISSVLEKGERLPOPICTIDVYIMVYKMMIDSECRPRELVEFSFMAADPORFV 821
QY 988 VIQ-NEDIGPASPDLSTYRSLLEDGMDGLVDABEYLYVPOGFCPPDPAPAGGVNHR 1046
DB 822 VIQGERHMLPSPPTSKFYRTLMEEBDMEDIVDADEYLVPHOGFF----- 866
QY 1047 HRSSSTRGSGDGLTLGLEPSEEARSPPLAPSEGASDVDFDGLMGAKGLQSLPTDHP 1106
DB 867 -NSPST-----SRTPLSLSLSTSN-----NSATKCIDRNGH-- 898

```

```

QY 1107 SPLQRYSEDPVLPPESTDGVAPLTCSPQPEYVNOBVPPOPSPSPREGFLPAPAPAGAT 1166
DB 899 -----PREDFGL-----PAPELVNQ-LMPKRPSTAMONQIYNYISLT 936
QY 1167 -IERAKTSPGXGVVAVFAGGAVENPEYL 1197
DB 937 AISKLPMSDRYQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
ID 085468;
AC 085468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian erythroblastosis virus (7s34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8217326; PubMed=2897102;
RA Scotting P., Vennartom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CA0F8AF4 CRC64;

Query Match 24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1,2e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPBACDVCACAHYKDPFCVACRCPGVPDLISYPIKFPDEBACQPCPINTHSCVDL 637
DB 1 GP--DHCMKAHFIDGPHCVKACAPAGVLGENDTL-VKYNADANAQCQLCHPCTRCCKP 57
QY 638 DDKGCRAEORASPLTSIVSAV-V-GILVVVLGVVFGILIKRQOKIRKRYTMRLLDET 636
DB 58 GLEGC---NGSKTPSIAGVVGGLCLVVGIGIGLYLRR-HIVKRTLRLLOEREL 113
QY 697 VEPLTSPGAMPNOAMRILKETELRKVYLGSGAFGVVYKGINIPDGENVYKIPVAIKVLR 756
DB 114 VEPLTSPGAMPNOAHLRIKETEPKVKVLFQAFGVYKGLMIPDEKXTIPVAIKELR 173
QY 757 ENTSPANKKEILDEAYVAVGVSPPVSRLLGICLTSTVOLTLQMPYGLLDHVENRGR 816
DB 174 EATSPANKKEILDEAYVAVSVDPNPHVCRLLGICLTSTVOLTLQMPYGLLDHVENRGR 233
QY 817 IGSODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHYKITDFGLAQLDIDE 876
DB 234 IGSQYLLNMCQIAKGMNYLEERHMYHRDLAARNVLKTPQHVKITDFGLAQLDIDE 293
QY 877 YHADGKVPKMMALLESILRRFTHOSDWSYGVYVWELMTFGAKPRYDGIIPAREIPDL 936
DB 294 YHAEKGKVPKMMALLESILHRIYTHOSDWSYGVYVWELMTFGSKPRYDGIIPAREISSVLE 353
QY 937 KGERLPOPICTIDVYIMVYKMMIDSECRPRELVEFSFMAADPORFVVIQ-NEDIG 995
DB 354 KGERLPOPICTIDVYIMVYKMMIDSECRPRELVEFSFMAADPORFVVIQDERNH 413

```

QY 996 PASPLDSTYRSLLEDDMDGLVDAAEYLVPOGFCPPDAPGAGGVHHRHSSSTRSG 1055  
DB 414 LBSPTDSKFXRTLMEDMEDVIDADDEVLYPHGCF-----NSPST---454  
QY 1056 GGDLTGLERSEEARSP-----APSEAGSDVFDGDLGMAAKGLQSLPHPHDSPLQ 1110  
DB 455 -----SRTPLSSLSATSNNSATNCIDRNG-----H-----481  
QY 1111 RYSEDFVPLPSEFDGVAPLTCSPQPEYVNOQDVPRQPSPREGPLPARPAGAT-LER 1169  
DB 482 -----PAREDFL-----PAPRYNO--LHPKPKSTAMVQIQIYINISLRISK 523  
QY 1170 AKTLSPKKGKGVKDVFAFGAVENPEYL 1197  
DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

RESULT 15  
Q9WVF5 PRELIMINARY; PRT; 655 AA.  
ID Q9WVF5  
AC Q9WVF5;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Epidermal growth factor receptor (epidermal growth factor receptor isoform 3).  
GN EGFR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=LIVER;  
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maizle N.J.;  
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAJ;  
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maizle N.J.;  
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr1 transcripts encoding truncated receptor isoforms."  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=LIVER;  
MEDLINE=21085560; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schiml L.M., Stabli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.D., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL: AF124513; AAD44149.1; -  
DR EMBL: AF275366; AAG28047.1; -  
DR EMBL: AF275364; AAG28047.1; JOINED.  
DR EMBL: AF275365; AAG28047.1; JOINED.  
DR EMBL: AK004944; BAB23688.1; -  
DR EMBL: AK004883; BAB23641.1; -  
DR EMBL: AK004911; BAB23662.1; -  
DR MGI: 95294; Egfr.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR002174; Furin-like.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR SMART: SM00261; FU\_3.  
KW RECEPTOR.  
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.5%; Score 1533.5; DB 11; Length 655;  
Best Local Similarity 44.4%; Pred. No. 4,1e-107;  
Matches 284; Conservative 99; Mismatches 233; Indels 23; Gaps 7;

QY 11 LILALPPGAA-STGYCTGDMKRLPASPEHLDMLRHYOGCGVQGNLELYLPTN 68  
DB 14 LITLCAAGGALAEKKVCGTSNRLTQGFEDHFLSGMYNNCEVLGLNLEITYQVN 73  
QY 69 ASLSEFDIOEVOGYLTAHNOVROVPLRLRYRGTOLFEDNYALAVDNGPLNNTTP 128  
DB 74 YDLSFKTTIQEAVGYLALNTERLPLENLQIRGNALYENTYALALISN----- 124  
QY 129 VTGASPGGLRELQRLSTELKGVLIQRNPOLCYQDTLMKDI-----FKKNNQLALTLI 184  
DB 125 -YGTNRGLRELPMRLQELIGAVFSPNNPILCNMDITQMRIVQNVFNSMSMDL--- 180  
QY 185 DTRNSRACHPCSPMGKSRGSGSSSDQSLRTYVAGGA-RCKGPLPTDCHEQCAAG 243  
DB 181 -QSHPSCKPKDPSCPNGSCMGSGEENCQKLTICAOQSHRCGRSPDCCCHNCACG 239  
QY 244 CNGPKSDCLACHFPHSGICELHCPALTYNTDTESHPNREGRTTFGASCTACPRNY 303  
DB 240 CTGPRESCLVQCFQDEATCTKPTPLMLYNPTTYQMDNPGKYSFGATCVKCKPRNY 299  
QY 304 LSTDVSCCTLVCPILHNOEYTAEDGTORCEKSKPCARVCYGLMEHLREVAVTSANIOE 363  
DB 300 VTDHSGCYRACQPDYVE-EDDGIRKCKKCDQPCRVKNGSIGEFKDTLSINATNIXH 358  
QY 364 FAGCKKIFGSLAFLEPSFDGDPASNTAPLOEQLOVETLEETLYGYLISAMPDSLPLDS 423  
DB 359 FKYCTAISGDLHLIPVAFGDSFTTRPLDPRELEIKTYKEITGFLLIQAWPDNMTDLH 418  
QY 424 VQONLOVITGRILNCAVYLTLOGLSIWGLRSLRELQSLALIHNNTHLCVHTVPWD 483  
DB 419 AFENLEIIRGRKXKHQGFSLAVAGLNTSLGRSLKEISDGVILISGNRLCYANTINWK 478  
QY 484 QLFRRPHQALHTANPREDECVGEGLAGHQLCARGHCWGGPQCNCQSGFLRGOCVEE 543  
DB 479 KLFGRPNQTKTMANRAEDCCAVNHVCPPLSSBECGWGEPDQVCSQNVSGRCVEK 538  
QY 544 CRYLQGLPREVYNARHCLPHECPQNGSVTCFGEADQCVACAHYKPPFCVAPCPSPG 603  
DB 539 CNLIEGPREFEIVNSICICHCHECLPQANNITCTGKGPNCITQCAHYIDGPHVKTCPCAG 598  
QY 604 VNPDLSTYPIWKFPRDEGACQCPINCTHSCVDLDDKGC 642  
DB 599 IGENNTL-VWKYADANNVCHLCHANCTGACAGPGLQGC 636

Search completed: July 22, 2003, 08:11:27  
Job time : 133.375 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds  
(without alignments)  
5267.077 Million cell updates/sec

Title: SEQ4-465-479-14

Perfect score: 6842  
Sequence: 1 MEALALCRMGILLALLPPGA.....TFKGTPTAENPEVLGLDVPV 1261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR\_73:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6626	96.8	1255	1 A24571	protein-tyrosine k
2	5846.5	85.5	1254	2 148161	p-185 precursor -
3	5846	85.4	1260	1 TYRTNU	protein-tyrosine k
4	3118	45.6	1210	1 GQHUE	epidermal growth f
5	3086	45.1	1210	2 A53183	epidermal growth f
6	3070.5	44.9	1223	1 TVCHLV	epidermal growth f
7	2942.5	43.0	1308	2 A47253	epidermal growth f
8	2661	38.9	1166	1 S06142	protein-tyrosine k
9	2381.5	34.8	1342	2 A36223	kinase-related tra
10	2300.5	33.6	1339	2 UC4387	epidermal growth f
11	1766.5	25.8	698	1 TVFVLV	protein-tyrosine k
12	1703	24.9	604	1 TVYUHH	protein-tyrosine k
13	1647	24.1	544	2 S35745	protein-tyrosine k
14	1640	24.0	545	2 S00727	kinase-related tra
15	1626.5	23.8	1330	1 GQFPE	epidermal growth f
16	1623	23.7	540	1 B44776	protein-tyrosine k
17	1621	23.7	540	1 TVFVBB	epidermal growth f
18	1478	21.6	644	2 A36325	epidermal growth f
19	1276	18.6	1323	2 E88257	protein-tyrosine k
20	1276	18.6	1374	2 S70712	protein-tyrosine k
21	1194	17.5	1369	2 S70713	protein-tyrosine k
22	1167	17.1	1717	1 A45558	epidermal growth f
23	1108	16.2	527	2 A42032	epidermal growth f
24	943.5	13.8	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	717	10.5	1363	2 T43220	insulin-like growt
28	703	10.3	1382	1 INHUR	insulin receptor p
29	702	10.3	1383	2 A36080	insulin receptor p

30	701.5	10.3	1372	2 A34157	insulin receptor p
31	691	10.1	1477	2 T18534	protein-tyrosine k
32	683.5	10.0	1300	2 A36502	insulin receptor-r
33	680	9.9	1607	2 T43212	insulin-like growt
34	671	9.8	1268	2 B36502	insulin receptor-r
35	645	9.4	1367	1 IGHUR1	insulin-like growt
36	632.5	9.2	2148	1 A56081	insulin receptor -
37	630	9.1	1390	2 T30346	insulin-like growt
38	625	9.1	1371	2 A33837	insulin receptor -
39	623	9.1	2101	2 S57245	insulin-like growt
40	601	8.8	987	2 A54092	insulin receptor (
41	597.5	8.7	952	2 I50612	protein-tyrosine k
42	597.5	8.7	984	2 A39753	protein-tyrosine k
43	587.5	8.6	977	2 S49004	tyrosine kinase Mp
44	586	8.6	1091	2 S33596	protein-tyrosine k
45	585.5	8.6	976	2 A36355	protein-tyrosine k

## ALIGNMENTS

## RESULT 1

A24571  
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human  
N.Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein ei  
C.Species: Homo sapiens (man)  
C.Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999  
C.Accession: A24571; A25491; A44188; B44188; I59509; I57622  
R.Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.  
Nature 319, 230-234, 1986  
A.Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth  
A.Reference number: A24571; MUID:86118663; PMID:3003577  
A.Accession: A24571  
A.Molecule type: mRNA  
A.Residues: 1-1255 <IAM>  
A.Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198  
R.Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985  
A.Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid  
A.Reference number: A25491; MUID:86016729; PMID:2995967  
A.Accession: A25491  
A.Molecule type: DNA  
A.Residues: 737-1031 <SEM>  
A.Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282  
R.Cousens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,  
Science 230, 1132-1139, 1985  
A.Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chr  
A.Reference number: A44188; MUID:86070181; PMID:2999974  
A.Accession: A44188  
A.Molecule type: DNA  
A.Residues: 740-910 <COU>  
A.Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989  
A.Accession: B44188  
A.Molecule type: mRNA  
A.Residues: 1-517, 'RALI', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>  
A.Cross-references: GB:M11730; NID:G183986  
R.King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985  
A.Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A.Reference number: I59509; MUID:85272597; PMID:2992089  
A.Accession: I59509  
A.Status: translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 832-909 <REX>  
A.Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808  
R.Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987  
A.Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcrip  
A.Reference number: I57622; MUID:87286898; PMID:3039351  
A.Accession: I57622  
A.Status: translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AAA58637.1; PID:9553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERBB2; NGU; NEU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 inase  
 F:1-21/Domain: signal sequence #status predicted <Sig>  
 F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:22-653/Domain: extracellular #status predicted <EXT>  
 F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:654-675/Domain: transmembrane #status predicted <TM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:68,124,187,253,571,629/Binding site: carboxylate (Asn) (covalent) #status predict  
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:753/Active site: Lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.8%; Score 6626; DB 1; Length 1255;  
 Best Local Similarity 97.1%; Pred. No. 2.1e-263;  
 Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;

QY 1 MELALCRWGLLALPLPAASTOVCTGDMKRLPASPETHLDMLRHLVYOGQVYQNL 60  
 DB 1 MELALCRWGLLALPLPAASTOVCTGDMKRLPASPETHLDMLRHLVYOGQVYQNL 60

QY 61 ELYLPPTNASTSFLDIOEVQGVLIANNOVQVPLQLRLIVRGTLQFEDNYALAVLNDG 120  
 DB 61 ELYLPPTNASTSFLDIOEVQGVLIANNOVQVPLQLRLIVRGTLQFEDNYALAVLNDG 120

QY 121 DPLNNTPTVTGASPGGLRELQSLTELKGVLIQORNPOLCYQDTILMKDIFHNKNOLA 180  
 DB 121 DPLNNTPTVTGASPGGLRELQSLTELKGVLIQORNPOLCYQDTILMKDIFHNKNOLA 180

QY 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRITVCAAGCARKCKPLPTDCHEQC 240  
 DB 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRITVCAAGCARKCKPLPTDCHEQC 240

QY 241 AAGCTGPRKSDCLACHFNHSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300  
 DB 241 AAGCTGPRKSDCLACHFNHSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPRLNHOEYTAEGTORCEKSPCARVCYGLCMOYIKANSKRIGIT 360  
 DB 301 YNYLSTDVGSCTLVCPRLNHOEYTAEGTORCEKSPCARVCYGLCMOYIKANSKRIGIT 360

QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVETLEETITGLYISAMPDSL 420  
 DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVETLEETITGLYISAMPDSL 420

QY 421 DLSVQNLQVTRIGRLIANGAYSLTIOGLISWGLRSRLRELGSSGLALIHNNHLCFVHTV 480  
 DB 421 DLSVQNLQVTRIGRLIANGAYSLTIOGLISWGLRSRLRELGSSGLALIHNNHLCFVHTV 480

QY 481 PMDOLFERNFTVSMFLRPKVSASHLECEVGEGLACHQLCARHGCMGSPGQCNVCSQF 540  
 DB 481 PMDOLFERNFTVSMFLRPKVSASHLECEVGEGLACHQLCARHGCMGSPGQCNVCSQF 540

QY 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECQONSVTGFGADQCVACAHYKDP 600  
 DB 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECQONSVTGFGADQCVACAHYKDP 600

QY 601 FCVACRPSGVKPDLSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAEGRASPSTSI 660  
 DB 601 FCVACRPSGVKPDLSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAEGRASPSTSI 660

DB 595 FCVACRPSGVKPDLSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAEGRASPSTSI 654  
 QY 661 VAAVVGILLVWVLGVVFGILIKRROOKIKRTYMRLLQSTELVEPLTPSGAMNOQMRI 720  
 DB 655 ISAVWGIILLVWVLGVVFGILIKRROOKIKRTYMRLLQSTELVEPLTPSGAMNOQMRI 714

QY 721 LKETELRKVKVLSGAFGVYKGIWIPDEBNYKIPAIKVLRENTSPKANKELUDEAYVM 780  
 DB 715 LKETELRKVKVLSGAFGVYKGIWIPDEBNYKIPAIKVLRENTSPKANKELUDEAYVM 774

QY 781 AGVGSPPYVRLIGICLTSIVOLVDTQMPYGCLLDHVENRGRIGSODLNMCMQIAKMS 840  
 DB 775 AGVGSPPYVRLIGICLTSIVOLVDTQMPYGCLLDHVENRGRIGSODLNMCMQIAKMS 834

QY 841 YLEDVRLVNRDLAARVNLVKSPPHVKITPFGRLALDIDETEHADGKVPIMMALESI 900  
 DB 835 YLEDVRLVNRDLAARVNLVKSPPHVKITPFGRLALDIDETEHADGKVPIMMALESI 894

QY 901 LRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMI 960  
 DB 895 LRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMI 954

QY 961 MYKCMWIDECRPRFRELVSFSSRMARDPQRFVITQNEIDLGPASPLDSTFYRSLLEDDDM 1020  
 DB 955 MYKCMWIDECRPRFRELVSFSSRMARDPQRFVITQNEIDLGPASPLDSTFYRSLLEDDDM 1014

QY 1021 GDLVDAEELVLPQGFPCFCDPAPGACGMVHHNRSSSTSGGDLTLGLBPSSEEA PRSP 1080  
 DB 1015 GDLVDAEELVLPQGFPCFCDPAPGACGMVHHNRSSSTSGGDLTLGLBPSSEEA PRSP 1074

QY 1081 LAPSEGASDVDPDGLGMGAAGLQSLPHHDSPLQRYSEDDPVPPLPSETDGVAPLTC 1140  
 DB 1075 LAPSEGASDVDPDGLGMGAAGLQSLPHHDSPLQRYSEDDPVPPLPSETDGVAPLTC 1134

QY 1141 POREYVNOBPVPOPPSPREGPLPAARAGATLERAKTSPGNGGVKDVAFGAVENP 1200  
 DB 1135 POREYVNOBPVPOPPSPREGPLPAARAGATLERAKTSPGNGGVKDVAFGAVENP 1194

QY 1201 EYLTPOGGAAPOPHPPPASPAFDNLVYDODPPERGAPSTFKGTPTJENBEYGLDVP 1260  
 DB 1195 EYLTPOGGAAPOPHPPPASPAFDNLVYDODPPERGAPSTFKGTPTJENBEYGLDVP 1254

QY 1261 V 1261  
 DB 1255 V 1255

RESULT 2  
 148161  
 p-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
 C:Accession: 148161  
 R:Nakamura, T.; Ushijima, T.; Iehizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Iehikav  
 Gene 140, 251-255, 1994  
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Reference number: 148161; MID:94193007; PMID:7908275  
 A:Accession: 148161  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <R5>  
 A:Cross-references: GB:D16295; NID:9493236; PIDN:BA003801.1; PID:9747595  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.5%; Score 5846.5; DB 2; Length 1254;  
 Best Local Similarity 85.3%; Pred. No. 1.4e-231;  
 Matches 1080; Conservative 60; Mismatches 109; Indels 17; Gaps 3;



```

QY      1 MELAALCRWGLLALLPFGAASSTOVCTGTDMKRLPASPEHLDMLHLYGCGVOVGNL 60
DB      1 MELAAMCGWGLLALLPFGASGTQVCTGTDMKRLPASPEHLDMLHLYGCGVOVGNL 60
QY      61 ELTYLPTNASTLSFLODIQEVQGYVLIAHNOVROVPLRLIRVGTQOLFEDNYALAVLDNG 120
DB      61 ELTYLPANASTLSFLODIQEVQGYVLIASOVRAVPLRLIRVGTQOLFEDNYALAVLDNR 120
QY      121 DPLNNTPTVYGASPGGLRELOLRSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNOLA 180
DB      121 DPLDNVTATGRTPEGLRELOLRSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNOLA 180
QY      181 LTLIDTRSRACHPCSPMKGSRGWGSESDCQSLTRTVCAAGCARCKPLPTDCCHQC 240
DB      181 LTLIDTRSRACHPCSPMKGSRGWGSESDCQSLTRTVCAAGCARCKPLPTDCCHQC 240
QY      241 AAGCTGPKHSDCLACLFHNSGICELCHPALVTNTPTFESMPNREGYTFGASCVTACP 300
DB      241 AAGCTGPKHSDCLACLFHNSGICELCHPALVTNTPTFESMPNREGYTFGASCVTACP 300
QY      301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTORCEKSKPCARVCYGLGMQYIKANSKFTIGT 360
DB      301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTORCEKSKPCARVCYGLGMQYIKANSKFTIGT 360
QY      361 ELEPAGCKKIFGSLAFIPESPDGDPASNTAPLOEOLQVFTTEITGYLYISAMPDLP 420
DB      361 IOEFAGCKKIFGSLAFIPESPDGDPASNTAPLOEOLQVFTTEITGYLYISAMPDLP 420
QY      421 DLSVFOULQVIRGRIHNGAYSLTLOGLISWLGRLSLRELGSGLALIHNNTHLCFPHTV 480
DB      421 DLSVFOULQVIRGRIHNGAYSLTLOGLISWLGRLSLRELGSGLALIHNNTHLCFPHTV 480
QY      481 PMDOLFRFNNFTVSMFLRVKVSASHL-----EECVBEGSLACHOLCARGHCWGPGPTQCV 535
DB      481 PMDOLF-----RNPHQALLHSGNPSEBEGCLKDFACVFLCAHGHGCMGSGPTQCV 529
QY      536 NCQQLFARGQCEVEECRLQGLPREYVNAHRLCPHPECOPONGSVTTPGPPADCCVACAH 595
DB      536 NCQQLFARGQCEVEECRLQGLPREYVNAHRLCPHPECOPONGSVTTPGPPADCCVACAH 595
QY      596 YKDPFVCVARGCPGKPDLSYMPIMKFPDEEGACOPCPINCTHSCVLDLDDGCPABORAS 655
DB      596 YKDPFVCVARGCPGKPDLSYMPIMKFPDEEGACOPCPINCTHSCVLDLDDGCPABORAS 655
QY      656 PLTSIVASAVGILLVVLGVVFGILLIKRROQKIRKYMRLLOETELVEPLTPSGAMPNQ 715
DB      656 PLTSIVASAVGILLVVLGVVFGILLIKRROQKIRKYMRLLOETELVEPLTPSGAMPNQ 715
QY      716 AQWRILKETELRKVKVYSGAFGVYVYGIWIPDGENVKIPVAIKVLRRENTSPKANKELD 775
DB      716 AQWRILKETELRKVKVYSGAFGVYVYGIWIPDGENVKIPVAIKVLRRENTSPKANKELD 775
QY      776 EAYVMAGVSPYVSRLLGICLTSTVOLVTOMLPYGCLLDHYRENRGLSGSODLLNMCMOI 835
DB      776 EAYVMAGVSPYVSRLLGICLTSTVOLVTOMLPYGCLLDHYRENRGLSGSODLLNMCMOI 835
QY      836 AKGMSYLEDLVRLVHRDLAARNVLYKSPNHVKITDFGLARLLIDIDETEHADGKVPKMM 895
DB      836 AKGMSYLEDLVRLVHRDLAARNVLYKSPNHVKITDFGLARLLIDIDETEHADGKVPKMM 895
QY      896 ALESILRRRTTHOSDWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLEPPICTI 955
DB      896 ALESILRRRTTHOSDWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLEPPICTI 955
QY      956 DVMYIMVCKMIMISECPRFRELVSFSPARMADPQRFVIVONEDLGASPLDSFFYISL 1009
DB      956 DVMYIMVCKMIMISECPRFRELVSFSPARMADPQRFVIVONEDLGASPLDSFFYISL 1009
QY      1016 EDDMDGLVDABEYLVPOQGFCEPDPAFGAGVNHRRSSSTRSGGDLTLGLEPSEEE 1075
DB      1016 EDDMDGLVDABEYLVPOQGFCEPDPAFGAGVNHRRSSSTRSGGDLTLGLEPSEEE 1075
QY      1076 APRSPLAPSEGASDVFDGLGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVA 1135
DB      1076 APRSPLAPSEGASDVFDGLGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVA 1135

```

```

DB      1070 PPRSPLAPSEGASDVFEGLMGATKGPQISPRDLSPLQRYSEDPVLPSETDGYVA 1129
QY      1136 PLTCSPOPEYNOBVDVPOPPSPREGPLPAARPGATLERAKTISPGKNGVYKDVFAFGG 1195
DB      1130 PLACSPQPEYNOBVDVPOPPSPREGPLPAARPGATLERAKTISPGKNGVYKDVFTTGG 1189
QY      1196 AVENPEYLTPOGAAPPHPPAPSPAFDNLVYWDOPPERGAPSTFKGPTAENPEYL 1255
DB      1190 AVENPEYLVPRGSGASQPH-PPALCPAFDNLVYWDOPPERGAPSTFKGPTAENPEYL 1248
QY      1256 GLDVPV 1261
DB      1249 GLDVPV 1254

RESULT 3
TNRITU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C/Accession: A24562; A61204
R/Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A/Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A/Reference number: A24562; MUID:66118662; PMID:3945311
A/Accession: A24562
A/Molecule type: mRNA
A/Residues: 1-1260 <BAR>
A/Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
R/Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen
Carcinogenesis 12, 1975-1978, 1991
A/Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
2-thiazolyl]formamide or N-methyl-N-nitrosourea
A/Reference number: A61204; MUID:92035293; PMID:1682063
A/Accession: A61204
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 637-663 'V', 665-702 <MAS>
A/Note: authors translated the codon GCA for residue 25 as Val
C/Genetics:
A/Gene: neu
C/Superfamily: epidermal growth factor receptor; protein kinase homology
C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-19/Domin: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-660/Domin: transmembrane #status predicted <TMN>
F:733-988/Domin: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: lys #status predicted
F:982,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.4%; Score 5846; DB 1; Length 1260;
Best Local Similarity 85.8%; Pred. No. 1,4e-231;
Matches 1084; Conservative 52; Mismatches 118; Indels 10; Gaps 3;

QY      1 MELAALCRWGLLALLPFGAASSTOVCTGTDMKRLPASPEHLDMLHLYGCGVOVGNL 60
DB      1 MELAAMCGWGLLALLPFGASGTQVCTGTDMKRLPASPEHLDMLHLYGCGVOVGNL 63
QY      61 ELTYLPTNASTLSFLODIQEVQGYVLIAHNOVROVPLRLIRVGTQOLFEDNYALAVLDNG 120
DB      64 ELTYLPANASTLSFLODIQEVQGYVLIASOVRAVPLRLIRVGTQOLFEDNYALAVLDNR 123
QY      121 DPLNNTPTVY-GASPGGLRELOLRSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNOLA 179
DB      124 DPODNVAASRTPGRTPEGLRELOLRSLTEILKGVLIORNPOLCYQDNVLMKDVFRKNNOLA 183
QY      180 ALTLIDTRSRACHPCSPMKGSRGWGSESDCQSLTRTVCAAGCARCKGRLPTDCCHQC 239
DB      184 APVDIDTRSRACHPCAPACKCKDNHCGSESDCQILTGITCTGCAACCKGRLPTDCCHQC 243

```

240 CAAGCTGPKHSDCLACLAHFHNSGICELHCPALVTYNTDTPESMPNPEGRTYFGASCYTAC 299  
 244 CAAGCTGPKHSDCLACLAHFHNSGICELHCPALVTYNTDTPESMPNPEGRTYFGASCYTAC 303  
 300 PYNALSTDVSGCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLNQYIKANSKFIGI 359  
 304 PYNALSTEVGSCCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLNQYIKANSKFIGI 363  
 360 TELEFAGCKKIFGSLAFLPESPDGPASNTAPLOEQOVPFTLEITGYLYTSMWPSL 419  
 364 NVQGFDEGCKKIFGSLAFLPESPDGPASNTAPLOEQOVPFTLEITGYLYTSMWPSL 423  
 420 PDLVFOQLQVIRGRILHNGAYSLTLQGLISWLSRLSRLSGSLALIHNTHLCPYHT 479  
 424 RDLVFOQLRIRGRILHNGAYSLTLQGLISWLSRLSRLSGSLALIHNTHLCPYHT 483  
 480 VPMDOLEFNNFTVSWFLRVPKVASHLBE--CVSEGLAQHQLCARGHCWGPFTQVCNC 537  
 484 VPMDOLEFNNFTVSWFLRVPKVASHLBE--CVSEGLAQHQLCARGHCWGPFTQVCNC 536  
 538 SGPRLGCEVCECRVLOGLPREYVNAHCLPCHPCOPONGSVTCGPEADOCVCAHYK 597  
 537 SHFLNGCEVCECRVWKGLPREYVSDKCLPCHPCOPONSETGFGSEADQCAACAHYK 596  
 598 DPPECVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPL 657  
 597 DSSCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPL 656  
 658 TSIVSAVVGILLVVVAVFGLILKRRQOKIRKYMRRLLQETELVEPLTPSGAMPNOAQ 717  
 657 TFIATVAGVGLFLVLVVVAVFGLILKRRQOKIRKYMRRLLQETELVEPLTPSGAMPNOAQ 716  
 718 MRILKETELRKVKYLGSGAFGVVYKGIWIPGSEANKIVAKIVLEKNTSPKANKIILDEA 777  
 717 MRILKETELRKVKYLGSGAFGVVYKGIWIPGSEANKIVAKIVLEKNTSPKANKIILDEA 776  
 778 YVMAGVSPYVSRLLGICLTSTVQVLTQMPYGLLDHVRNKGRLSGODLIMCMQIAX 837  
 777 YVMAGVSPYVSRLLGICLTSTVQVLTQMPYGLLDHVRNKGRLSGODLIMCMQIAX 836  
 838 GMSYLEVRLVHRDLAARNVLYKSPNHVYKIDFGIARLLDIDETEHADGCKVPIKXMAL 897  
 837 GMSYLEVRLVHRDLAARNVLYKSPNHVYKIDFGIARLLDIDETEHADGCKVPIKXMAL 896  
 898 ESILRRRTTHQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLKSGRLPQPPICITDV 957  
 897 ESILRRRTTHQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLKSGRLPQPPICITDV 956  
 958 YMINVKCMWIDSECRPRELVSERFARMARDPQRFVIONEDLGSPASPLDSTFYSLLED 1017  
 957 YMINVKCMWIDSECRPRELVSERFARMARDPQRFVIONEDLGSPASPLDSTFYSLLED 1016  
 1018 DDMGDLVDAEYLYPQGFPCPDPAFGCGMHHRRSSSTRSGGDLTLGLEPSEEPAP 1077  
 1017 DDMGDLVDAEYLYPQGFPCPDPAFGCGMHHRRSSSTRSGGDLTLGLEPSEEPAP 1076  
 1078 RSPLAPSGAGSDYVDGDLGWAAGKAGSLPPTHDSPLORYSEDPVTLPSFTDGYVAPL 1137  
 1077 RSPLAPSGAGSDYVDGDLGWAAGKAGSLPPTHDSPLORYSEDPVTLPSFTDGYVAPL 1136  
 1138 TCSPQPEYVNOQPPVRRPQSPRECEGLPAPAPAGATLERAKTLPQKNGVNVDFAFGAV 1197  
 1137 ACSPQPEYVNOQPPVRRPQSPRECEGLPAPAPAGATLERAKTLPQKNGVNVDFAFGAV 1196  
 1198 ENPEYLTPOGGAQAQHPHPAPAFSPAFDNLTYWDQPPERGAAPSPFTKGTPTAENPEYIGL 1257  
 1197 ENPEYLTPOGGAQAQHPHPAPAFSPAFDNLTYWDQPPERGAAPSPFTKGTPTAENPEYIGL 1256  
 1258 DVPV 1261  
 1257 DVPV 1260

RESULT 4  
 GORUE  
 epidermal growth factor receptor precursor - human  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999  
 C:Accession: A00641; A25772; S30024; A36672; A0642; A43615; A23062; A05281; A60143; A  
 R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.;  
 rg, P.H.  
 Nature 309, 410-425, 1984  
 A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression  
 A:Reference number: A00641; MUID:84219729; PMID:6328312  
 A:Accession: A00641  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <U>L>  
 A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924  
 A:Note: the authors translated the codon AAG for residue 540 as Aaa  
 R:Rishi, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985  
 A:Title: Characterization and sequence of the promoter region of the human epidermal g  
 A:Reference number: A25772; MUID:85270438; PMID:2991899  
 A:Accession: A25772  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-29 <TSH>  
 A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272  
 R:Hailey, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.  
 Oncogene Res. 1, 375-396, 1987  
 A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification  
 A:Reference number: S30024; MUID:88217333; PMID:3329716  
 A:Accession: S30024  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HA2>  
 A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119  
 R:Hailey, J.D.; Waterfield, M.D.  
 J. Biol. Chem. 266, 1746-1753, 1991  
 A:Title: Contributory effects of de Novo transcription and premature transcript termin  
 A:Reference number: A36672; MUID:91107677; PMID:1988448  
 A:Accession: A36672  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HAL>  
 A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271  
 R:Hailey, J.D.; Waterfield, M.D.  
 J. Biol. Chem. 266, 1746-1753, 1991  
 A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN  
 A:Reference number: A00642; MUID:84245835; PMID:6330563  
 A:Accession: A00642  
 A:Molecule type: mRNA  
 A:Residues: 'RCAMRA', 150-187, 'KSUQNAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-3  
 , 798-799, 'ND', 802-811, 'R', 813-942 <XUY>  
 A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF rec  
 R:Lin, C.R.; Chen, W.S.; Krutger, W.; Stolarky, L.S.; Weber, W.; Evans, R.M.; Verma,  
 Science 224, 843-848, 1984  
 A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificatio  
 A:Reference number: A43615; MUID:84196372; PMID:6328261  
 A:Accession: A43615  
 A:Molecule type: mRNA  
 A:Residues: 713-964 <LIN>  
 A:Experimental source: epidermoid carcinoma cell line A431  
 R:Simmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.  
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
 A:Reference number: A23062; MUID:85046483; PMID:6093780  
 A:Accession: A23062  
 A:Molecule type: mRNA  
 A:Residues: 1028-1210 <SIM>  
 R:Weber, W.; Gill, G.N.; Speiss, J.  
 Science 224, 294-297, 1984  
 A:Reference number: A05281; MUID:84172183; PMID:6324343  
 A:Accession: A05281  
 A:Molecule type: protein  
 A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R.Russo, M.W., Lukka, T.J.; Cohen, S.; Staros, J.V.  
J. Biol. Chem. 260, 5205-5208, 1985  
A>Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor  
A.Reference number: A60143; MUID:85182650; PMID:2985580  
A.Accession: A60143  
A.Molecule type: protein  
A.Residues: 740-744, 'X', 746-747 <RUS>  
R.Mroczkowski, B., Mosig, G.; Cohen, S.  
Nature 309, 270-273, 1984  
A>Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase  
A.Reference number: A38023; MUID:84191554; PMID:6325948  
A.Content: annotation; receptor activity  
A.Note: The EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
R.Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
Cell 59, 33-43, 1989  
A>Title: Functional independence of the epidermal growth factor receptor from a domain I  
A.Reference number: A33331; MUID:9003223; PMID:2790960  
A.Contents: annotation; internalization signal  
C.Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex  
C.Genetics: -  
A.Gene: GDB:EGFR  
A.Cross-references: GDB:120610; OMIM:131550  
A.Map position: 7p12.3-7p12.1  
C.Superfamily: epidermal growth factor receptor; protein kinase homology  
C.Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase  
F.1-24/Domain: signal sequence #status predicted <SIG>  
F.15-1210/Product: EGF receptor #status predicted <KAT>  
F.725-645/Domain: extracellular #status predicted <EXT>  
F.7-300/Domain: EGF receptor extracellular domain repeat <EEL>  
F.390-600/Domain: EGF receptor extracellular domain repeat <EEZ>  
F.646-668/Domain: transmembrane #status predicted <TM>  
F.669-1210/Domain: intracellular #status predicted <INT>  
F.710-975/Domain: protein kinase homology <KIN>  
F.718-726/Region: protein kinase ATP-binding motif  
F.999-1046/Region: coated-pit mediated internalization signal  
F.1047-1210/Region: inhibitory  
F.128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F.745/Active site: Lys #status experimental

Query Match 45.6%; Score 3118; DB 1; Length 1210;  
Best Local Similarity 49.5%; Pred. No. 3e-120;  
Matches 630; Conservative 178; Mismatches 349; Indels 116; Gaps 25;

Oy	11	LILLALPPGAA--STGYCTGTDMKLRLPASPETHLIDMLRHLYOGCGVVGNIETLYLPNTN	68
Dd	14	LALLALCPASRALREKKVKCCGGTSNNLTQLGTFEDHFLSLQRMNFENCEVLGNLEITVQGN	73
Oy	69	ASLSFLDDLOEVGVGVYLIANNOVRQVPLQRLRIYRGTLDFBNVALAVLDNGDPLANTTP	128
Dd	74	YLSPFKTTIOEVAGVYLIANTVERIPLENQIIRGNMYENSAVALAVLSND-----	126
Oy	129	VTCGAPGLAREILOLRSLTEILKGVLIIQRNPOLCYODITILMWDIRPHKNQGLALVIDNR	188
Dd	127	---ANKTGLKEPLMRNLQELIHGAIVRSNNPALCNVESIQWRDIVSSDPISMMSMDPQN	183
Oy	189	SFACHPFCSPMCKSGRCWGESSEDCQSILRTVCAGCA-RCKGPLPTDCCHQCAGACTGP	247
Dd	184	IGSCCKCDPSCPMGSCWGAGEENCQKLTLLICAQCCSGRCKGSKGPSDCCHNCAGACTGP	243
Oy	248	KHSDDLACLHPNNSGICELHCPLLVYNTDTFESMPNPEGGYTRGASCYTACPYYLSTD	307
Dd	244	RSDCLVCKKFRDEATCKDTCPLMLYNPTTYQMDVNPBEGKYSPFATCVCKKCPRYYVVVD	303
Oy	308	VGSCTLVLCPLHNQVEATADGTORCEKSKPCARCVYGLGMQYIKANSKFITGLE-FAG	366
Dd	304	HGSCVRAAGADSYEM-EEDGVRRCKCKGECGRCKYCONGIQGEIFK-DLSISANTNIKRPN	361
Oy	367	CKKITGSLAFLESFGDDPASNTAPLOPOLQVFETLEEITGYLYISAMPDSLPLDSVQ	426
Dd	362	CTSIGGDTHILPVAFRGDSFTHTPPDLPOEIDLKTAYKEITGFLTIQAMPENRTDLHFE	421
Oy	427	NLOVIRGRILNHGAVSLTLOGLGISWLGLSLRELISGLALIHHNTHLCPFNHYWDOLF	486

Db 422 NLEIIGRTQHQOFLSAVLSNLTSLGLSLAKLSIGDVIISGKMLCYANTIMWKLF 481

Qy 487 RFNNFVFSFWLRAVPKVSASHLE-ECVGEGLACHOLCARGHCMGPRPTOCVNCSCQFLRGOE 545

Db 482 GTSG-----QTKTKIISNRGENSKATGQVCHALCSPEGCMGPRPCVSCRNVSRGHE 534

Qy 546 CEEECVLOGLPREYVNAHHCLEPCPCOPONGSVTCFGEPAQCVACAHYKDPCCVAR 605

Db 535 CYDKCKLLEGEPRPEFVENSECICHPRECLPQAMNITCTGRGPDNCTICCAHYIDGPHCVKT 594

Qy 606 CPBGVPRDLSYMPIMKPRDEEGACOPCPINCTHSCVDLDDGCAPEAGASPLTSIVSANV 665

Db 595 CPAGVGENNTLL-VMKYADAGVCHLCHPNCCTGCTGPGLECGPFGKPIR--SIATGCV 651

Qy 666 G--ILLVVLGVVFGILLKRRQOKIRKTYMRLLQETELVEPLTPSGAMPQAOAMRLK 722

Db 652 GALLLLLLVVALGIG---LFMRHRHIVAKRLTRLLQERLVEPLTPSGEAPQALLRLIK 708

Qy 723 ETEBLRVKVLGSGAFGTVYKGINIPDGENVKIPVAIKVLRENTSPPKANKIIDEAYVMAG 782

Db 709 ETEFKKIKVLGSGAFGTVYKGLWIPGEKVKYIPVAIKELREATSPKAKEIIDEAYVMAS 768

Qy 783 VQSPVYSLIGLITLSTVOLTOLMPYGCILLDHNREKRGRLGSDLLMWQOIAAGMSTL 842

Db 769 VDPHPHCRLLGLICLITSTVOLTOLMPGCLLDYREKMDNGSYLLNMVQOIAAGMNTL 828

Qy 843 EDVRLVHRDLAARNVLVKSPNNHKTDFGLARLLDIDETETHAAGKVPIMMALESTLR 902

Db 829 EDRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEETHAEGKVPIMMALESTLH 888

Qy 903 RRFTHQSDVMSYGVTVWEIMLTFGAKPYDGIIPAREIPDLLEKGRLPPOPCITIDVYIMV 962

Db 889 RIYTHQSDVMSYGVTVWEIMLTFGSKPYDGIIPASISILEKGERLPPOPCITIDVYIMV 948

Qy 963 KCMMTIDSECRPRRELVSERSMARDPQRVVVIO-NDLCPASPLDSTFFRSLLEDMDG 1021

Db 949 KCMMTIDADSRPKRRELIFESKMARDPQRVLVIOGDERMLPSTPDSNFYALMDEEDMD 1008

Qy 1022 DLVDAEEVLYVPOGFCPPDPAPAGCGVHHRHSSSTRSGGDLTLGLEPSEEAAPRSL 1081

Db 1009 DVVDADEVYIIPGGFF-----SSPSTSRPL 1034

Qy 1082 APSEGAQDVFPDGLMGAAKGLQSLPTHDPSPLORYSEDPTVLPSET--DGYVAPLTC 1139

Db 1035 ISSLATSNN--NSTVACIDRNGLQSCPIKEDSFQORYSSDPTGALTEDSIDTDL----- 1087

Qy 1140 SPQPEYVQNPQVROPQPSREGPLPAARPRGATLIERAKTISPCNGVYKQVFAFGAVEN 1199

Db 1088 -VPEEYING-SVPKRAGQVNPVYHNPQNP-----APSRDDPHYOD--PHSTAGN 1135

Qy 1200 PEYLL-TPQGAAPQPHRAPFASPAFNLVYWDQ-----DP-----PEHGAPST 1242

Db 1136 PEYLLTVQ-----PTCVNSTFDSPAHMAQKSHQISLDNPDYQDDFFPKAKKNGI 1186

Qy 1243 FKGTPTAENPEYLL 1255

Db 1187 FKGS-TAEVAYEYLL 1198

RESULT 5

A53183

epidermal growth factor receptor precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999

C:Accession: A53183, A43818, S24842, A28941, S65325, I49643

R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Eard, H.S.; Jenkins, N.A. Genes Dev. 8, 399-413, 1994

A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor

A:Reference number: A53183, MUID:94170986; PMID:8125255

A:Accession: A53183

A:Molecule type: mRNA

A:Residues: 1-1210 <LUE>

A:Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
Oncogene 6, 673-676, 1991  
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site  
A:Reference number: A43818; MUID:91232666; PMID:2030916  
A:Accession: A43818  
A:Molecule type: mRNA  
A:Residues: 1-714 <AVI>  
A:Cross-references: GB:X59698  
R:Elisinger, D.P.; Serreio, G.  
Submitted to the EMBL Data Library, June 1992  
A:Reference number: S24942  
A:Accession: S24942  
A:Molecule type: mRNA  
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>  
A:Cross-references: EMBL:Z12608  
R:Heisermann, G.J.; Gill, G.N.  
J. Biol. Chem. 263, 13152-13158, 1988  
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated  
A:Reference number: A28941; MUID:88330814; PMID:3138233  
A:Accession: A28941  
A:Molecule type: protein  
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,  
R:Hibbe, M.L.; Dunn, A.R.; Alexander, W.S.  
Submitted to the EMBL Data Library, April 1994  
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor  
A:Reference number: S45325  
A:Accession: S45325  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-971, 'K', 973-1210 <VER>  
A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CA55587.1; PID:9488831  
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.  
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993  
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse  
A:Reference number: I49643; MUID:9126380; PMID:7678348  
A:Accession: I49643  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 12-20, 22-132 <RES>  
A:Cross-references: GB:I06864; NID:9193001; PIDN:AA53029.1; PID:g567201  
A:Gene: EGFR  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:648-670/Domain: transmembrane #status predicted <TM>  
F:712-977/Domain: protein kinase homology <KIN>  
F:720-728/Region: protein kinase ATP-binding motif  
F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental  
F:697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental  
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental  
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental  
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.1%; Score 3086; DB 2; Length 1210;  
Best Local Similarity 49.4%; Pred. No. 6e-119;  
Matches 632; Conservative 170; Mismatches 358; Indels 120; Gaps 27;

QY 11 LILALPFGAA--STVCTGTDMKLRASPETHLMDLRHYGCCOVVQNLLETLYPTN 68  
DB 14 LILALCAAGALEBKVKCGQSTSNRLTOLGTFEDHFLSLQRMVNNCEVALNLETYYORN 73  
QY 69 ASLSFLDIDQEVQGVYLIANNOVROVPLQRLRIYRGTLQFEDNYALAVLNDGPIANTTP 128  
DB 74 YDLSPFKTIQEVAGVYLIANTVERIRPLENLIQIRGNALVENTYALALISN----- 124  
QY 129 VTGASGCGARELQRLSTETLKGVYLIQRPOLCYOTIIMKDI---FKNQALATLI 184  
DB 125 -YGTNTRJRELRLMRLQELILIGAVRPSNPIICNMDDIQRDIVQNVFWSNMSMD--- 180  
QY 185 DTRSRACHPSCPMCKSGRCMGSSSEDCSLTRTVACGGA-RCKGFLPDCCHQCGAAG 243  
DB 181 -QSHPSRCRKCDSPCRGSCMGSGEENQCULTKIICAQCCSHRCRGRSPSDCHQCGAAG 239

QY 244 CTGPKSDCLACLPNHSIGICELCPALVYNTDTFSSMPNREGRYFGASCYTCAPVNY 303  
DB 240 CTGRRESUCLVCQRFQDEBATCTCPMLNPTTYQNDVNBEGRTSGATCVKCKCPVNY 299  
QY 304 LSTDVGSCTVCPPLHNOVTAEDESTORCEKSCPKCARCYVGLGMOYIKANSKFIGITILE 363  
DB 300 VVTDHSGCVARCGDYIEV-EEDGIRCKKCKDDGCRVACNIGIIGERF-DLSTINATIK 357  
QY 364 -FACCKKIFGSLAFLPSPFDGDPASNTAPLOPBOUVFETLEBITGVYIYISAMPDPLD 422  
DB 358 HFKYCTAISGDLHLIPAFKGDSTFRTPLDPRLEILIKTYKEITGFFLIQAMPDPMWDL 417  
QY 423 SVFQNLQVINGRIILHNGVSLTLQGLISWLGRLSRLBELSGGLALIHNTHLCPHTYPM 482  
DB 418 HAFENLSTIRKTKQHOQPSLAVVGLNITSGLSLKEISGDVLIISGNRLCYANTINW 477  
QY 483 DQLFRPNFTVSEFWLRVPKVSASHLE-ECVDEGLACHQLCARGHGCGPQCVCNCSQFL 541  
DB 478 KKLFGFTN-----QKTKIMNNAEKDCRVNHVNCPLCSSECGKPREPDCSCNVVS 530  
QY 542 RQECVEECRYLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPF 601  
DB 531 RGRCEVCKCNILBEGPREFVENSECICHPCECLPQAMNITCTGRGPNICICAHYIDGP 590  
QY 602 CVARCPGVPRDLSYMPIWKPDEGACQCPICHTSCVDLDKCGPABGRASPITSIV 661  
DB 591 CVKTCFPAIGENNTL-VMKRADANNVCHLNACTYGCAGPGLOGCEWPSGKPIBSIA 649  
QY 662 SAVGILLVVLGVVFGI-LIKRQOKIRKTYMRLLOETELVEPLTPSGAMPQAOQRI 720  
DB 650 TGIYGLGLFIV-VALDIGLFMRRIHVRKTRLRLLQERELVEPLTPSGAPQOAHRI 708  
QY 721 LKETELKRVYLSGAFGTYYKGIWIPDENVKIPIVAILRENTSPKANEILDEAVYM 780  
DB 709 LKETEPFKIKIVLGSAGFGTYVKGIMPEGEVKIPIVAILRENTSPKANEILDEAVYM 768  
QY 781 AGVSPVSRLLGLCLSTVOLVQLMPYCGCLDHYENGRGLSODLLNMCQIACGMS 840  
DB 769 ASVDNPRVCRLLGICLSTVOLVQLMPYCGCLDHYENHNDIGSQVLYNNCQIACGMN 828  
QY 841 YLEVDRLVHRDLAARNLVKSPNHVKITDPGLARLIDETERYADGCKVPIKMALESI 900  
DB 829 YLEDRRLVHRDLAARNLVKTPQVKTITDGLALGLAEKEYNAEGCKVPIKMALESI 888  
QY 901 LRRRTQSGVMSYGVTVWELMTFGAKPYGCIIPAREIPDLIEKGERLPQPICTIDVYMI 960  
DB 889 LHRITYHQSDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPICTIDVYMI 948  
QY 961 MVKCMIMDSRCRPRPRELVEEFSRMAPDPRFVYIC-NEDLGRASPLDSTFYSLEDD 1019  
DB 949 MVKCMIMDADSRPRFRELIEFSOMARDPRVYLTQGDERRMHLPSPTDSNPFYRALMDEED 1008  
QY 1020 MGDVLVAEYLVPOQGFCEPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEEARPS 1079  
DB 1009 MEDVVADEYLIIQOGF-----NSST-----SRT 1034  
QY 1080 PLAPSBGASDVDFDGLGMAKAGLSLPTHDSPLORYSEDPVPLPSET-DGYVAPL 1137  
DB 1035 PLISSLATSNN-----NSTVACINRNGSCRYKEDAFIORYSSDPGAVTEBNDIDAFI--- 1087  
QY 1138 TCSPOPEYVQNPVRQPRPREGLPRAARADATLEBATISPGKNGVYKDVFAFGAV 1197  
DB 1088 ---PVEEYVQ-SVPRKPRAGSVQNPVHNPRLP-----APGRDLAYON--PHSNAY 1133  
QY 1198 ENPEYL-TPGGAAPQPHPRPAPSPAFDNLVYWDQ-----DP-----PERGAP 1240  
DB 1134 GNPBYLNTAQ-----PTCLSSGPNFALMIQSGHMSLDNDYQDDFFPKETKPN 1184  
QY 1241 STFKGTPTAENPPLYGLDVP 1260  
DB 1185 GIFKPG-PTAENAEYLAVARP 1203

## RESULT 6

TVCHLV

epiderma: growth factor receptor precursor - chicken

N:Contins: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Gallus gallus (chicken)

C&gt;Date: 28-Feb-1986 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000

C/Accession: A27720; A00643

R/Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart

Mol. Cell. Biol. 8, 1970-1978, 1988

A/Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A/Reference number: A27720; MUID:88261272; PMID:3260329

A/Accession: A27720

A/Molecule type: mRNA

A/Residues: 1-1223 &lt;LAX&gt;

A/Cross-references: GB:M20386

R/Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M

Cell 41, 719-726, 1995

A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro

A/Reference number: A00643; MUID:85228222; PMID:2988784

A/Accession: A00643

A/Molecule type: mRNA

A/Residues: 585-1223 &lt;NLL&gt;

A/Cross-references: GB:M10066

C/Genetics:

A/Genes: erbB

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor

specific protein kinase

F/1-30/Domain: signal sequence #status predicted &lt;SIG&gt;

F/31-1223/Product: epidermal growth factor receptor #status predicted &lt;MAT&gt;

F/31-654/Domain: extracellular #status predicted &lt;EXT&gt;

F/81-307/Domain: EGF receptor extracellular domain repeat &lt;EE1&gt;

F/397-610/Domain: EGF receptor extracellular domain repeat &lt;EE2&gt;

F/655-677/Domain: transmembrane #status predicted &lt;TM&gt;

F/678-1223/Domain: intracellular #status predicted &lt;INT&gt;

F/719-984/Domain: protein kinase homology &lt;KIN&gt;

F/727-735/Region: protein kinase ATP-binding motif

F/136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #

F/192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted

F/687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F/754/Active site: Lys #status predicted

F/1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.9%; Score 3070.5; DB 1; Length 1223;

Best Local Similarity 48.2%; Pred. No. 2.6e-118;

Matches 629; Conservative 175; Mismatches 347; Indels 153; Gaps 28;

QY 8 RMGLLALPFGAA-----STOVCTGTMKLRPLPASPETHLDMRLHYGCGQVVGNGLE 61  
 DB 13 RGAALVLLGLGVALCSAVEEKVCOGTNNKLTQLGHVEDHFTSLQRMNNCEVVLNLE 72  
 QY 62 LTYLPTNASSLFDIOGVGVYLIANOVROVPLQRLTVRGQLFEDNYALAVLNGD 121  
 DB 73 IIVENHRDITFTKIOEVAGVYLIANMVDVIPLENLQIIRGVNLVDSFALAIVSNYH 132  
 QY 122 PLNNTTPVTGASPGRLQELQRLSLTELKGVYLIQRNPOLCYODTILMKOIFPHNQAL 181  
 DB 133 -NMKTQ-----GARELPMKRLSELINGVYKISNNPKLCMDTVAMNDIITSRK-PL 182  
 QY 182 TLID-TNRBACHPCSPMCKGSRGWESSSDCSLRTVAGCA-GCKPLPTDCHEQ 239  
 DB 183 TVLDFASNLSSCKCHPNCTEDHCWGAGEONCQTLTKVICAQCCSGRCGRKVPSDCCNHQ 242  
 QY 240 CAAGCTGPKHSDCLACHPHFHSIGICELHCPALVTYNTDFEPMNPBGRRTFGASCYTAC 299  
 DB 243 CAAGCTGPRSDCLACKRFDDATCTKOTCPPLVLYNFTTYQMUNPEGKYSFATYATREC 302  
 QY 300 PNYLSTDVGSCTLVCPRLHNOETADGTQRCCKSKPCARVCYGLQMOYIKANSKFIGI 359  
 DB 303 PHNYVTVTDGSGCVSRGNTDTYER-EENGVRKCKKCGGLGSKVNGIGIGLKIILS-INA 360  
 QY 360 TELE-PAGCKKIFGSLAFPLPESFDGPASNTAPLOEBOLOVFTLEBTITGYLYISAMPDS 418

DB 361 TNIDSFKNCTKINDVSLIPVAFGLDAFTKTLPLDPKKLDVFRVKEISGFLLIQAMPDN 420  
 QY 419 LPDLSVONLOVIRGRILHNGAYSLTQGLGISWGLSRSLAEISGLLIHNNTHLCVH 478  
 DB 421 ATDLYAFENLTIYGRKQKQOISLAVNLKISGLSLISDGIAlMKNNKLYCYAD 480  
 QY 479 TVPMDOLFERNPTVFSFMLRVPKVSASHLECEVGLACHOLCARGHGCGPPTOCVNC 538  
 DB 481 TNNRSLFATQS-----QTKKIIONNNKNDCTADRHVCBDPLCGDVGCMGCPFFHCFSR 534  
 QY 539 QFLRGQCEVCEKVLQGLPREYVNRHCLPCHPECCQNG--SVTCFGRPADQVCACAH 595  
 DB 535 FFSRQKCEVQCNLTQCEPREFERSKCLPCHSECLVONSTAYNTTCSGPPDHCMKCAH 594  
 QY 596 YKDPFPVAPRPSGVKNDLSMPYWKRPDEGAQOPCPINCTHSCVDLDDKCAEAOAS 655  
 DB 595 FIDPFPVAPRPSGVKNDLSMPYWKRPDEGAQOPCPINCTHSCVDLDDKCAEAOAS 655  
 QY 656 PLTISVAVV-GILLVVVLGVVFGILLIKRQOKIRKTYMRRLQETLEVEPLTPSGAMPN 714  
 DB 651 KTSPIAGVGVGLCLVVGIGLGLYLRH-HYRKRTLRLDLERLVEPLTSGEAPN 709  
 QY 715 QAOIRILKTELARKVYVLSGAFGTYYKGIWIPDGENVKIPVAIKVRENTSPKANKEL 774  
 DB 710 QAHRLIKETEFKKVYVLSGAFGTYYKGIWIPDGENVKIPVAIKVRENTSPKANKEL 769  
 QY 775 DEAVTMAGVSPYSLRILGICLTSTVOLVQLMRYGCLLDHVENRGRGLSGODLNMCMQ 834  
 DB 770 DEAVTMAGVSPYSLRILGICLTSTVOLVQLMRYGCLLDHVENRGRGLSGODLNMCMQ 829  
 QY 835 IAKMSYLEDVRLVHRLAARNVYVKSPPNHYKIDFGRLDIDETEVHDDGGVPIK 894  
 DB 830 IAKMNTLEERLVRHRLAARNVYVKSPPNHYKIDFGRLDIDETEVHDDGGVPIK 889  
 QY 895 MALESILRRPFTQSDVMSYGVTVWELMTFCAKPYDIPAREIDLEKGRLEPPPICT 954  
 DB 890 MALESILRRPFTQSDVMSYGVTVWELMTFCAKPYDIPAREIDLEKGRLEPPPICT 949  
 QY 955 IDVYMTWKCMVIDSEKRPREFVSEFSKMARPOGFVYQ-NEDIGAPSLDSTFYRS 1013  
 DB 950 IDVYMTWKCMVIDSEKRPREFVSEFSKMARPOGFVYQ-NEDIGAPSLDSTFYRS 1009  
 QY 1014 LLEDDBDGLVDAEYLVVPOQGFPCPDPAAGAGMVAHRRSSSTRGGDLTLGLESE 1073  
 DB 1010 LMEBEDMEDVDAEYLVVPOQGFPCPDPAAGAGMVAHRRSSSTRGGDLTLGLESE 1038  
 QY 1074 EEARPSPL-----APSEAGSDVDFDGLGMAAGKLSLPTHDSPLQRYSEDPVPLPS 1128  
 DB 1039 ---SRTPLSLSLATSNSNATNCID-----RNGGHPVREDSFYQVSSDPTGNFLE 1087  
 QY 1129 ET--DGYVAPLFTCSPOPEYVNOVDVPPPSPREBGLPAARPAATLERAKTISPGKNGV 1186  
 DB 1088 ESIDDFGL-----PAPEYVNO--LMPKKS-----TAMVQNOI 1118  
 QY 1187 VKDYF-----AFGAVENREYVLTPOGGAAPHPPAFPAFNDLYWDO 1231  
 DB 1119 YNNISLTAISKLPMSDRYONSHSTAVDNPEYL-----NTQSGPLAKTVFESSPYWQ 1170  
 QY 1232 -----DPE-----RGAPSTFKGPTAENPEYLGIDV 1260  
 DB 1171 SGHQNINLNDPVDYQODPLPMTKPNGLIKYPALENPEYLVAAAP 1214

## RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C/Species: Homo sapiens (man)

C/Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999

C/Accession: A47253

R/Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A/Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderma







```

Db      238 PODTDFACAHFNDGACVPRCPPLVYNKLTFOLEBNPHTKYQYGVGVASCPHNFV-V 296
Qy      307 DVSGCTLVCPHLHNGEYVAEDGTORCEKCSKPCARVCGIGLMQYIKANSKF--IGITBLE- 363
Db      297 DQTS CVRACPPDKMEVD--KNGLKMCEPCGGICCPACCEGTG-----SSRQYTVDSNIDG 350
Qy      364 FACCCKIFGSLAFLPESFPDDPASNTAPLOPEOLQVETLEETLEITGYIYSAMPISLDLS 423
Db      351 FVACTKILGNLDFLITGLNDPMMHKIPALDEPKLVNRTYREITGYINISWPRPHMNF 410
Qy      424 VFONLOVIRGRILHNGAYS--LTLOGGIGISWLGRLSRLSGSLALIHNTLHLCFVHTVPW 482
Db      411 VFSNLTTIGGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRIYISANQOLCHHSLNW 470
Qy      483 DOLFERNFVFWLARP---KVSASH---LEECVNGSLACHOCARHGCMGPRPTCVN 536
Db      471 TKY-----LRGPTEERLDIKNRRPRDDVAAGKVCDFLCSSGGCMGPRGQCLS 519
Qy      537 CSQFLRGOECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCACAHY 596
Db      520 CRNYSRGVCVTNCHNPLNGEPREPAHEACGSCHEQPMEGTATCMGSSGSDTCAQCANH 579
Qy      597 KDRPFCAVCPGKVPDLSTWPIWKFPDEBACQPCPINCTHSCVDLDDKGCRAQRA-- 654
Db      580 RDRPHCVSSCPHGYLG--AKGPRIKYPDVONECRPCHENCTOGCKGPELDCLGQTLVLI 637
Qy      655 --SPLTISVAVVGIILVVVLGVFGLIKRROOKIR--KYTWRRLLQETELVEPLTPSGA 711
Db      638 GKTHLTALTATVING--LVVIFMMLGGFTLWGRRIQNKAMRRYLYRGSEIEPLDPS-E 694
Qy      712 MPQAQMRILKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVAIKVIRENTSPRANK 771
Db      695 KAKKVLARIKETEELRLKVLGSGGVFTGKGVWIPGESIKIPVCIKVEDSGRSGSQ 754
Qy      772 EILDEAVWAGVSGSPYSLIGLICTSTVOLVTOLMYGCCLLDHVENKRGSLSDILLN 831
Db      755 AVTDHMLAISLSHANIYRLGLCPGSSLDLVTOYELPGLSLDHDVORHGLGQILLNW 814
Qy      832 CMOIAKMSYLEDRVLRDLAARNVLSKSPNHKIDFGIARLDIDETEYHADGKVP 891
Db      815 GVQIAKMYLLEHGVNHRNLAARNVLLSKSPQOVADLPFGADLLPRDDQKQLYSEAKTP 874
Qy      892 IKKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAKEIPDLLEKGRLLPQP 951
Db      875 IKKMALESIHFGKYTHQSDVMSYGVTVWELMTFGAEFYAGRLAEVDDLEKGRLLAQ 934
Qy      952 ICTIDVYMWVKCMVIDSECRPRRELVSFSPMARDPORFVITQNEIDLPA--SPLDS 1008
Db      935 ICTIDVYMWVKCMVIDSECRPRRELVSFSPMARDPORFVITQNEIDLPA--SPLDS 993
Qy      1009 TFRSLLEDDMDGLVDAAEYLVPOQGFCCRDPARAGAGVNHNRHSSSTRSGGDLTLG 1068
Db      994 GLTNKLEVELEBELDLDLLEAEED-----NLATTLGASALSIP 1034
Qy      1069 LEP--SEEDARSPILABEGAGSDVFDGLCMGAKGLQSLPTHD--PSPLORYSEDPVPL 1126
Db      1035 VGTILNPRGQSILSPSSGY--MPMNQNLGESCOESAVSSSSERCPRVSLH-----PM 1087
Qy      1127 P-----SETDGYVA-----PLTSGQPR-----YVQNPVRVROPSPRRE 1160
Db      1088 PRGCLASSESGHVTGSEAELOEKVSMCRSRSRSPRPGDSAYHSQRLSLTPVPLS 1147
Qy      1161 GP-----LPAARPAATILERAKTLP--GKNGV-----KQVFAFGAVENREY 1202
Db      1148 PROLEBEDVNGYMPDTHLKTGTSRREGTILSSVGLSVLTBEEDD-----EYEV 1199
Qy      1203 LTPOGGAAPQPPRPPAPAFNLYWD-----QDPPERCARSTFKGTPTAA 1250
Db      1200 MNRRRHSP--PHPPRPSLEELGYEYWDVGSLSLSTGSCPLHVPIMPATGTTPDE 1258
Qy      1251 NPEYL 1255
Db      1259 DVEYM 1263

```

```

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N/Alternate names: ErbB3 protein; HER3 protein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C/Accession: J04387
R/Heliyer, N.O.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A/Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein
A/Reference number: J04387; MUID:96096535; PMID:8522190
A/Accession: J04387
A/Molecule type: mRNA
A/Residues: 1-1339 <HEL>
A/Cross-references: GB:U29339; NID:9915389; PID:9915390
A/Experimental source: liver
A/Note: The authors translated the codon AAG for residue 369 as Thr and GTT for residue
C/Comment: This protein is a functional heregulin receptor that transduces signals to
C/Genetic8:
A/Genes: ErbB3
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homo
C/Keywords: Atp; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #stratus predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #stratus predicted <MAT>
F:640-659/Domain: transmembrane #stratus predicted <TM>
F:705-970/Domain: protein kinase homolog <KIN>
F:713-721/Region: protein kinase Atp-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match      33.64; Score 2300.5; DB 2; Length 1339;
Best Local Similarity 40.18; Pred. No. 7.5e-87;
Matches 520; Conservative 171; Mismatches 430; Indels 175; Gaps 36;

Qy      3  LVALCRWGLLALIPGAA---STQVCTGDMKRLPASPETHDMLRHLHYOGCGVQGN 59
Db      7  LQVLC-----FLSLARSEBNGNSQAVCPGLNLGSLVGDADNOVQTLKYEKEEVWGN 62
Qy      60  LELTYLTNASSLFLQDIOEVQGYLIANHQVNPQLRLRYRGTLFEDNVALAVLDN 119
Db      63  LEIVLTGHNADLSFLQWIREVTAVLVANMEFSVLPRLRYVNGTVQVDOCKFAIPW-- 120
Qy      120  GDLNLTPTVYTGASPGGLRLQLRLSLTEILKGVTLQIRNQLCYQDTIILMKDIFHNK 179
Db      121  ---LNVNT---NSSHALRQLKFTQLTEILISGVYIEKNKLCIMDTIIDRDIYRVR-- 170
Qy      180  ALTLIDTNRGRACHPSCPMCKSGRCWGESEDCSLTRTYCAGGC--ARCKGPLPTDCHE 238
Db      171  GAEIVKKNANCPCHIEVKG--RCWGPGRDDCOILTKTICAPQCNRCGFPNQCCHD 229
Qy      239  QCAAGCTGPHNSDCLACLHFNHSGICEHLPAVYNTDTFESMNPREGRYTFGASCVTA 298
Db      230  ECAGGCGSGPDQTCFCARFNDGACVCRPREPLVYNKLTFOLEBNPHTKYQYGVGVAS 289
Qy      299  CPYNYLSTDVSGCTLVCPHLHNGEYVAEDGTORCEKCSKPCARVCGIGLMQYIKANS 356
Db      290  CPNHFV--VDQTFVCRACPPDKMEVD--KNGLKMCEPCGGICCPACCEGTGSSRQYTV 347
Qy      357  IGITLEFAGCKKIFGSLAFLPESFPDDPASNTAPLOPEOLQVETLEETLEITGYIYSAMP 416
Db      348  ID-----GFVACTKILGNLDFLITGLNDPMMHKIPALDEPKLVNRTYREITGYINISW 403
Qy      417  DSLPDLVSFONLOVIRGRILHNGAYS--LTLOGGIGISWLGRLSRLSGSLALIHNTLHLC 475
Db      404  PHMNFVSFNSNLTTIGGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRIYISANQOLC 463
Qy      476  FVHTVMDQLFRFNNFTVSWFLRVPKVSASHL-----ECVBEGLACHOLCARGRWCP 529
Db      464  YHSLNWTRL-----LRGPSEERLDIYDRPRLGCLAEGRKVCDFLCSSGGCMG 512
Qy      530  GPQCVNCSQFLRGOECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQ 589

```



```

Db      513 APOGCLSCRNYSREGVCTVHCNPLQGEPRFVHEAOFCSCHEPCLPMEGTSTYNGSGDA 572
Qy      590 CVACAHYKDPFPCVARGCPGPDLSYPMIKPPEDEGACOPCPINCHSG--VDLDDXG 647
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      573 GARCNAHRDGHPCNSCPHGLG--AKCPITKTPDAQNECPCHENCTOGNGPELQCL 630
Qy      648 CPAEQRASPLTISAVVGLLVVGLGVFGILIKRROOKIR-KYTMRLLOETELVEPL 706
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      631 GQAEVLMSKPHLVIAVTVG--LAVILMILGGSFLYMRGRRIQNBRAMRYLERGESIEPL 688
Qy      707 TPSCAMNQAOAMRLKTELAKVYLGSAFGTYKGIWIDGENVKIPVAIKVLRENTS 766
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      689 DPS-EKANQVLAIRIFKETELRKLKVLGSGVGTGVIHKGWIEGESIKIPVCIKYIEDKSG 747
Qy      767 PKANKELIDEAYVMAAGSPVYSRLIGICTSTVQLVTLQMPYGCILDHVENRGRLSGQ 826
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      748 ROSQAVTDHMLAVGSLDHAHIVRLIGLCPSSSIQLVLYPLSLDLHVAKHETLAPQ 807
Qy      827 DLLNMQOIAKMSYLEDLVLRDLAARVLYKSPNHVKITDFGLARLDIDETEVHAD 886
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      808 LLNMVQOIAKMGVYLEBHSWVHRDLARNVLMKSPSQVADLFPDDKQLHLS 867
Qy      887 GGAKPIMMALESLIRRFTHQSDVWSYGTVMELMTFGAKPYGIGIPAREIPDLLEKGER 946
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      868 EAKPIMMALESLIRRFTHQSDVWSYGTVMELMTFGAKPYGIGIPAREIPDLLEKGER 927
Qy      947 LPOPICTIDVYIMVCMIMIDSECRPFRELVESEFSRMARDPQRFVITQMEDIGPASP 1006
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      928 LAPOICTIDVYIMVCMIMIDENIRPFKELANEFIRMAIDPRYLVIKAS--GPGIP- 985
Qy      1007 DSTFFRSLDEDDMGDLVDAEYLVPQGFPCPDPAAGAGVHHRHRSSTRSGGDLT 1066
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      986 -PAEPFSLTTKEL-----QEAELEPEL-----DLD 1010
Qy      1067 LGLEPSE-----EAPRSPLASEG-----AGSDVVDGDLGM 1098
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1011 LDLEAESEGATSLGSLALSLPTGTLTRPRGSQLSPSSGYPMNQSLGACLDASVTLG 1070
Qy      1099 GAAKGLQSLPTHPDSPLORYSEDPVLPPESETDGV---APL-----TC-----SPQ 1142
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1071 GREQFSRPISLH-PIPRGR-----PASESSBCHVTGSEAELEKYSVCNRSRNSRPP 1122
Qy      1143 PE---VYNQPDVARPPSPREGR-----LPAPAPAGATLERAKTLASP-GKNGV 1186
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1123 PRGSAVSHQSRHSLTPTPLSPGLPEEDNGVYMPDTHLRGASSSREGTLSSVGLSSV 1182
Qy      1187 V-----KQVPAFGAVENPEYLLTPQGAAPQPHPP 1216
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1183 LGTEDEDED-----EYEVYMKRGRGSP-PRPP 1209

```

RESULT 11

TVPLV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C:Species: avian leukosis virus, ALV

C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999

C/Accession: B00643

R:Nilgen, T.W.; Matoney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M. Cell 41, 719-726, 1985

A>Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and processing

A/Reference number: A00643; MUID:85228222; PMID:2988784

A/Molecule type: mRNA

A/Residues: 1-698 <N1>

A/Cross-references: GB:M10066, GB:M13881, NID:g211749, PIDN:AAA48763.1, PID:g211750

A/Note: In Genbank entry CHKERB8F, release 109.0, the source is designated as Gallus gallus

C/Comment: This protein is synthesized as a gag-env-erbB protein.

C/Genetics:

A:Gene: gag-env-erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P

F:1-6/Product: gag protein (fragment) #status predicted <AG>

```

F:1-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/bomaim: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match      25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 2.5e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy      584 GPEADQCVACAHYKDPFPCVARGCPGPDLSYPMIKPPEDEGACOPCPINCHSGVDL 643
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      60 GP--DHGKCAHFIIDGPHCAKCPAGVLGENDTL-VMKYADANAVACQLCHPCTRGCKCP 116
Qy      644 DDKCPAEQRASPLTISAVV--GILVVLVGLGVFGILIKRROOKIRKTYMARLLQETEL 702
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      117 GLEGP---NGSKPTSLAAGVGLCLLVVGLGLGLYLRGR-HIVRRKTRRLRLQEREL 172
Qy      703 VEPLTPSCAMNQAOAMRLKTELAKVYLGSAFGTYKGIWIDGENVKIPVAIKVLR 762
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      173 VEPLTPSCAMNQAOAMRLKTELAKVYLGSAFGTYKGIWIDGENVKIPVAIKVLR 232
Qy      763 ENTSPKANKEILDEAYVMAAGSPVYSRLIGICTSTVQLVTLQMPYGCILDHVENRGRL 822
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      233 EATSPKANKEILDEAYVMAAGSPVYSRLIGICTSTVQLVTLQMPYGCILDHVENRGRL 292
Qy      823 LGSDDLNMQOIAKMSYLEDLVLRDLAARVLYKSPNHVKITDFGLARLDIDETEVHAD 882
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      293 IGSQYLLNMQOIAKMGVYLEBHSWVHRDLARNVLMKSPSQVADLFPDDKQLHLS 352
Qy      883 YHADGKVPIMMALESLIRRFTHQSDVWSYGTVMELMTFGAKPYGIGIPAREIPDLLE 942
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      353 YHADGKVPIMMALESLIRRFTHQSDVWSYGTVMELMTFGAKPYGIGIPAREIPDLLE 412
Qy      943 KGERLPQPICTIDVYIMVCMIMIDSECRPFRELVESEFSRMARDPQRFVITQMEDIGP 1001
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      413 KGERLPQPICTIDVYIMVCMIMIDSECRPFRELVESEFSRMARDPQRFVITQMEDIGP 472
Qy      1002 PASLPLDSTFFRSLDEDDMGDLVDAEYLVPQGFPCPDPAAGAGVHHRHRSSTRSG 1061
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      473 LPSPPLDSTFFRSLDEDDMGDLVDAEYLVPQGFPCPDPAAGAGVHHRHRSSTRSG 513
Qy      1062 GGDVTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAGLQSLPTHPSP 1116
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      514 -----SKTPLLSSLSATSNNSATNCTD-----RNGQGHVPRDESPVQ 550
Qy      1117 RYSEDPVLPPESET--DGYVAPLTCSPQPEYVNOVDVARPPSPREGLPAPAPAGATLE 1174
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      551 RYSEDPVLPPESET--DGYVAPLTCSPQPEYVNOVDVARPPSPREGLPAPAPAGATLE 585
Qy      1175 RAKTLSPGKGVKGVKQVDF-----AFGAVENPEYLLTPQGAAPQPHPPAF 1219
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      586 ----TAMVQNOIVNNISLTAISKLPMDSRQYNSHSTAVDNPEYL-----NTQSP 633
Qy      1220 SPAPDNLYYMDQ-----DPEE-----RGAPSPFFKTPPAENPEYLGDPV 1260
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      634 KTFEESPPMIQSGNHQINLNDPDYQDFLENETKPGGLKVPAAENPEYLRVAP 689

```

RESULT 12

TVYUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)

C:Species: avian erythroblastosis virus

C/Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999

C/Accession: A00644; A38022

R:Yamamoto, T.; Mashida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K. Cell 35, 71-78, 1983

A>Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fam

A/Reference number: A00644; MUID:84026539; PMID:6313229

A/Molecule type: DNA

A/Residues: 1-604 <YAM>

A/Cross-references: GB:K01216; NID:g209676; PIDN:AAA442400.1; PID:g209676

R, Debuitte, B.; Henry, C.; Benaïssa, M.; Biserre, G.; Claverie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1984  
 A:Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of A:Reference number: A38022; MUID:84223957; PMID:5328658  
 A:Accession: A38022  
 A:Molecule type: DNA  
 A:Residues: 1-28, 'F', 141-145, 'V', 147-152 <DEB>  
 A:Cross-references: GB:K02006  
 C:Genetics:  
 A:Gene: erba  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific F:130-395/Domain: protein kinase homology <KIN>  
 F:138-146/Region: protein kinase ATP-binding motif  
 F:165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;  
 Best Local Similarity 52.2%; Pred. No. 8.5e-63;  
 Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;  
 QY 593 CAHYKDPFCVACRPSGVKPDLSYMPWKFPDEBEGACOPCPINCTHSCVDLDDKCGCPAQ 652  
 DB 3 CAHFIDGPHCVKACGAVLGENDTL-VKRYADANVCQLCHPNCRGCKGPGLECGP--- 58  
 QY 653 RASPLTSIVSAV-V-GILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGA 711  
 DB 59 NGSKTPSIAGVGVGLCLVVGVLGILGLYLR-R-HIVKRTLRLLQREILVEPLTPSGE 117  
 QY 712 MPNOAMRIKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIVAEINTSPKANK 771  
 DB 118 APNOAHILKETEERKVKVLSGAFGIVKGLMPBGEKVIPIALIELEBASPANK 177  
 QY 772 EILDEAVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGCCLLDHVENRGLSGODLNM 831  
 DB 178 EILDEAVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGCCLLDYREHKDNGSYLLNM 237  
 QY 832 CMQIAKMSYLEDRVLRDLAARVLYKSPNHVKITTFGLARLIDDETEYHADGGKVP 891  
 DB 238 CVOIAKGMNLYEERLVLRDLAARVLYKTPQHVKITDFGLAKLGADEKHYHAEKGKVP 297  
 QY 892 IKMALESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGI-PAREIPDLLEKGERLP 951  
 DB 298 IKMALESILRRITTHOSDWSYGVTVWELMTFGSKPYDGI-PASEISSVLEKGERLP 357  
 QY 952 ICTIDVYIMVCKMWDISECRPRRELVSFSPMAADPQRFVIO-NEDLGPASPLDSTF 1010  
 DB 358 ICTIDVYIMVCKMWDIDSRPKFRELIAFSPKMARDPRLVIOGDERMHLPEPTDSKF 417  
 QY 1011 YRSLLEDDMDGLVDAEYLYVPOGFCFDPAPAGAGMHHHRSSSTRSGGDLTGL 1070  
 DB 418 YRSLMEEDMEDIVDAEYLYVPHOGFF-----NSPST----- 449  
 QY 1071 PSEEARSPPL-----APSEAGSDVFPDGLGMGAAGLQSLPTHPDRLQRYSEDPVP 1125  
 DB 450 -----SRPLSSLSLSTSNNSATNCID-----RKGQGHVPREDSFVQRYSDDPGN 495  
 QY 1126 LPSET-DGYVAPLTCSPQPEYVNOQPVRRPQSPRREGPLPARPAGATLERRAKTLSPGK 1183  
 DB 496 FLEESIDDGFL-----PAREYVNO--LMPKKPSTAM----- 524  
 QY 1184 NGVYKDVFAF-----GAVENPEYLTPOGGAAPQPPPPAPSPAFD 1224  
 DB 525 --VONOQYVNFISLTAISKLPMSDRYONSHSTAVDNPEYL-----NTNOSPPLAKTVFE 574  
 QY 1225 NLVYWDODPPERCAPSTFKCTPAENPEY 1254  
 DB 575 SSPYMIQSGNHQ-----INLDNPDY 594

RESULT 13  
 S35745  
 C:Species: avian erythroblastosis virus  
 C:Species: avian erythroblastosis virus

C:Date: 03-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 28-Feb-1997  
 C:Accession: S35745  
 R:Vennstrom, B.  
 Submitted to the EMBL Data Library, March 1993  
 A:Reference number: S35743  
 A:Accession: S35745  
 A:Molecule type: DNA  
 A:Residues: 1-544 <VEN>  
 A:Cross-references: EMBL:X12707  
 C:Genetics:  
 A:Gene: erba  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific F:135-400/Domain: protein kinase homology <KIN>  
 F:143-151/Region: protein kinase ATP-binding motif  
 F:170/Active site: Lys #status predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;  
 Best Local Similarity 54.9%; Pred. No. 1.5e-60;  
 Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;  
 QY 584 GPEADOCVCAHYKDPFCVACRPSGVKPDLSYMPWKFPDEBEGACOPCPINCTHSCVDL 643  
 DB 1 GP--DHCKCAHFIDGPHCVKACGAVLGENDTL-VKRYADANVCQLCHPNCRGCKGP 57  
 QY 644 DDKCPAQRASPLTSIVSAV-V-GILLVVVLGVVFGILIKRQOKIRKYMRLLOETEL 702  
 DB 58 GLECGP--NSKTPSIAGVGVGLCLVVGVLGILGLYLR-R-HIVKRTLRLLQREIL 113  
 QY 703 VEPLTPSGAMBNQAMRIKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKYL 762  
 DB 114 VEPLTPSGEAPNOAHLRIKETEFKVKVLSGAFGVYKGLMPBGEKVIPIALIELEB 173  
 QY 763 ENTSPKANKETLDEAVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGCCLLDHVENRGR 822  
 DB 174 EATSPKANKETLDEAVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGCCLLDYREHKDN 233  
 QY 823 LGSODLNMCMQIAKMSYLEDRVLRDLAARVLYKSPNHVKITTFGLARLIDDETE 882  
 DB 234 IGSQYLLNMCMQIAKGMNLYEERLVLRDLAARVLYKTPQHVKITDFGLAKLGADEKE 293  
 QY 883 YHADGKVPIMKMALESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGI-PAREIPDL 942  
 DB 294 YHAEKGKVPIMKMALESILRRITTHOSDWSYGVTVWELMTFGSKPYDGI-PASEISSV 353  
 QY 943 KGERLPQRPITIVYIMVCKMWDISECRPRRELVSFSPMAADPQRFVIO-NEDLG 1001  
 DB 354 KGERLPQRPICITIVYIMVCKMWDIDSRPKFRELIAFSPKMARDPRLVIOGDERMH 413  
 QY 1002 PASPLDSTFYRSLLEDDMDGLVDAEYLYVPOGFCFDPAPAGAGMHHHRSSSTRSG 1061  
 DB 414 LPSPFDSKFPRTLMEEDMEDIVDAEYLYVPHOGFF-----NSPST-- 454  
 QY 1062 GGDUTLGLSEEARSPPL-----APSEAGSDVFPDGLGMGAAGLQSLPTHPDPSLQ 1116  
 DB 455 -----SRPLSSLSLSTSNNSATNCIDRNG-----H----- 481  
 QY 1117 RYSEDPTVPPLPSETDGYVAPLTCSPQPEYVNOQPVRRPQSPRREGPLPARPAGAT-LER 1175  
 DB 482 -----FVREDGFL-----PAREYVNO--LMPKKSTAVONOYVYISLTAISK 523  
 QY 1176 AKTLSPGKNGVYKDVFAF-----GAVENPEYLTPOGGAAPQPPPPAPSPAFD 1203  
 DB 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14  
 S00727  
 C:Species: avian erythroblastosis virus  
 C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 28-Feb-1997  
 C:Accession: S00727  
 R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987  
 A>Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant  
 A:Reference number: S00727; MUID:88217326; PMID:2897102  
 A:Accession: S00727  
 A:Molecule type: DNA  
 A:Residues: 1-545 <SCO>  
 A:Cross-references: EMBL:X06943  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; phosphotransferase  
 F:135-400/Domain: protein kinase homology <KIN>  
 F:133-151/Region: protein kinase ATP-binding motif

Query Match 24.0%; Score 1640; DB 2; Length 545;  
 Best Local Similarity 54.9%; Pred. No. 2.9e-60;  
 Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 584 GPEADCVACAHYDPPFCVACRSPGKPDLSYPIKFPDEBAGQPCPINTHSCVDL 643  
 DB 1 GP--DHCMKCAHFIIDGPHCVACRAGVIGENDTL-VMKYADANAVCQLCHPCTRCKGP 57  
 QY 644 DDKCPABQRASPLTISVSAVV-GILVVVVGAVFGILIKRQOKIRKTYRRLLOETEL 702  
 DB 58 GLEBCP--NGSKTPSIAAGVVGGLCLVVGGLGILYLR--HIVRRKTRRLLOEREL 113  
 QY 703 VEPILPSGAMPNQOMRILKETELRKVVLGSGAFYVKGIMIPDGENVKIPVAIKYLR 762  
 DB 114 VEPILPSGEAPNQMHRLIKETEFKRVVLFQGFYVKGIMIPDGENVKIPVAIKYLR 173  
 QY 763 ENTSFKANKELDEAYVAGVSPYVSRLLGICLTSTVQVLTQMPYGCILLDHVRENKGR 822  
 DB 174 EATSPKANKELIDAYVMAVDNPHVCRLLGICLTSTVQVLTQMPYGCILLDIYREHKDN 233  
 QY 823 LGSODLLNMCQAIKMGSYLEDVRLVHRDLAARVNVKSPHNVKITDPSGLARLLDIDTE 882  
 DB 234 IGSQYLLNMCVQIAKMGVLEERHLVHRDLAARVNVKTPDVKITDPSGLAKOIGAEBKE 293  
 QY 883 YHADGKVPILKMALESILRRFTHSDVMSYGVTVWELMFGAKPYDGIPIAREIPDLLE 942  
 DB 294 YHAEQGVPIKMALESILHRIYTHQSDVMSYGVTVWELMFGSKPIDIGIASSETSVLE 353  
 QY 943 KGERLPDPPICTIDVYIMYKCMWIDSECRPREFVSEFSRMAPDPORFVVIQ-NEDLG 1001  
 DB 354 KGERLPDPPICTIDVYIMYKCMWSDSRPKFRELIAEFKMAPDPRLVVIQGDERRMH 413  
 QY 1002 PASPLDSTFYRSILEDMDGLVDAEYIVPQGFPCDPAPRAGAGMNHNRSSSTRSG 1061  
 DB 414 LPSPDTSKFRITLMEEDMEDIVDADEYLVPHQGF-----NSPST--- 454  
 QY 1062 GGDLLTGLPSEEEAPRSP-----APSEGASDVFPDGLQMGAKGLQSLPTIDPSPLO 1116  
 DB 455 -----SKTPLLSSLSATSNSNATCIDRNG-----H----- 481  
 QY 1117 RYSEDPVLPSETDGVAPLPCSPPEYVNOQDPVRQPSPREGRLPAARPAAGAT-LER 1175  
 DB 482 -----PYREDDGL-----PAPEYVNO--LMPKKRSTAMVQVIVYISLTATISK 523  
 QY 1176 AKTLSPKNGVNVDFAFGAVENPEYL 1203  
 DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

RESULT 15  
 GQFFE  
 epidermal growth factor receptor - fruit fly (Drosophila melanogaster)  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: Drosophila melanogaster  
 C:date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 11-Jun-1999  
 C:Accession: A00640; A38021  
 R:Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B. Z.  
 Cell 40, 599-607, 1985  
 A>Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding

A:Reference number: A00640; MUID:85124611; PMID:2982499  
 A:Accession: A00640  
 A:Molecule type: DNA  
 A:Residues: 1-1330 <LIV>  
 A:Cross-references: EMBL:K03054  
 R:Madsworth, S.C.; Vincent III, W.S.; Bildeau-Wentworth, D.  
 Nature 314, 178-180, 1985  
 A>Title: A Drosophila genomic sequence with homology to human epidermal growth factor  
 A:Reference number: A38021; MUID:85137938; PMID:2983232  
 A:Accession: A38021  
 A:Molecule type: DNA  
 A:Residues: 'A', 832-866, 'V', 868-943, 'QPSLVK' <WAD>  
 A:Cross-references: EMBL:X02293; NID:G7922; PIDN:CAA26157.1; PID:G929565  
 C:Comment: This sequence is tentative because the introns have not been identified.  
 C:Genetics:  
 A:Gene: FlyBase: Egfr  
 A:Cross-references: FlyBase: FBgn0003731  
 A:Map position: 2.57F  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase  
 F:1-732/Domain: extracellular #status predicted <EXT>  
 F:733-764/Domain: transmembrane #status predicted <TM>  
 F:765-1330/Domain: intracellular #status predicted <INT>  
 F:808-1072/Domain: intracellular #status predicted <INT>  
 F:816-824/Region: protein kinase ATP-binding motif  
 F:122-300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:843/Active site: Lys #status predicted  
 F:181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted

Query Match 23.8%; Score 1626.5; DB 1; Length 1330;  
 Best Local Similarity 30.2%; Pred. No. 2.4e-59;  
 Matches 418; Conservative 179; Mismatches 411; Indels 377; Gaps 45;

QY 80 VQGYVLAHNVQROVPLRLRIVRGTOF-----EDNYALVLDNGDPLNTPVTGASR 134  
 DB 38 ITNYIVIGLIDLPCTLSYRLQIRGRITFLSVSEBEKALVF-----TY 81  
 QY 135 GGLRELQRLSTELLKGVLIQRNPOLCYOTILMKDIFHKQNALALITDNRBACHP 194  
 DB 82 SKMTLETPDLRDVNLGVGVHNNYMLCHMTQISELVNSGTDAYVYVDFAPAREECPK 141  
 QY 195 CSPMKGSRGWGESESDQSLTRTVACAGCA--RCKGPLPTDCCHEOGACGCGKHSDC 252  
 DB 142 CHESTG--CMGEPRKQCKSKLTCSPOCGRGKYGKPECCCLFACAGCTGPTQKDC 200  
 QY 253 LACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCYACPVNYLSTVGSCT 312  
 DB 201 IACNPFDEAVSKCECPMRKYNPPTYVLETPNPSGKYAGATVCECP--CHILRNGACV 259  
 QY 313 LVCPYLNQEVTAEDGTCRCEKSKPCARVCYGLGMQYIKANSKFIGITEL-----EPAG 366  
 DB 260 RSCPQDKDKDQGE-----CVPCNGPCPKTC-----PGVTVLAHAGNIDSPRN 300  
 QY 367 CKKIFGSLAFPLESFDG--DPASNTA-----PLQRPQLOVEFTELEETGYLYISAMWS 418  
 DB 301 CTVIDGNIRILDQFFSGFDVYANYTNGPRYIPLDPEKRVFSTVKEITGYLNEGTHPQ 360  
 QY 419 LPDLVFPQNLQVIRGRILHNGAY-SLTLOGIGISMLGRLSRLREAGSLALIHNTHTLGFV 477  
 DB 361 FRNLSTFYRNLETIHGRQLMSMFALAIIVSSLSLEMRNKQISSGSVVIQHRDLCTV 420  
 QY 478 HTVPW-----DOLFRNNFTVS----- 494  
 DB 421 SNIRMPAIQKEPQKVWVNNLRADLGKFTLLISVOHNIIMHIFAIAREKMHLLGSV 480  
 QY 495 -----FMRLV-----PKYSA 504  
 DB 481 QRGRLGASWGSVPYLOELOFQNHRLRLMLYIOVINSITODKSNHQLTDACYSPSVPT 540  
 QY 505 S-----HLEE-----C-----VG 512  
 DB 541 SLTERARVYIQSAGLAMELOITAREASMRHSKTLRABGRQVRWVFLGVCSABAGIA 600

Qy	513	GLA-----CHOLCARHGMCGPPOCVNCSQPLFGQEVCECRVLQGLPREYV---N	562
Db	601	EPLGRAVCRKCHPLCELCTNYGHNBOYCSKCTHYKAREOCETC-----PADHYDEE	654
Qy	563	ARHCLPCHPECOPONGSVTCFGRPEADOCVACAHYK-----DPEF-----CVARCPSG	609
Db	655	QRECFQRHPEC---NG---CTGPGADDCCKSRNFKLFDANETGYVNSYTMPCNCKSPLE	708
Qy	610	VK-PDLSTMPYWKPFDERGACQPCPINCSTHSCVLDLDDKCPAEORASPLSYSAVVGIL	668
Db	709	MRHNVQYOTAGPY-----CAASPRSSKITANLD-----VNMIFIIIGAV	749
Qy	669	LVVVLGVVFGI-LYKRROOKIRKXT--MRLLJOETELVEPLTPSGAMPNOAMLIKETE	725
Db	750	LVPLTICLCVVTYICRQOKAKKETVKTMTALSGRESEPELRPSNIGANLCKRIIVDAE	809
Qy	726	LKRVKVLGSAFGTVYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILBEAYVMAGVS	785
Db	810	LKRGVGLMGAFGRYKGVWPEGENYKIPVAIKELKSTGAESEPTREAYIMASEEH	869
Qy	786	PYVRLLGICLTSTVOLVTOQLMPYGCILLDHYRENRGRSGSDDLNMWCQIATKMSYEDV	845
Db	870	VNLKLELAVCMSSQMMILTQMLPGCLLDYVRNNRDKSGSALLNWSTQIATKMSYEEK	929
Qy	846	RLVHRDLAARVNLVYK---SPNHVKITDEGLARLIDLIDETEXYHADGKVPYKMALESTLR	902
Db	930	RLVHRDLAARVNLVRLAGEDH----DEGLAKLSSDSNEKYKAAGKGFYKMALEECTRN	985
Qy	903	RRFTHQSDVMSYGVTTWELMTFFGAKPYDGPAREIPDLLEKGERLPOPPICTIDVYMW	962
Db	986	RVFTSKSYVMVAFGTIWEILLTFQGORPHENIPAKDIPDLIEVGLKLEQPEICSIDICYCTL	1045
Qy	963	KCMWIDSECRPRELVESEFRMARDPRFVJQNEDLG--PASPDLSTFRSLLED--	1018
Db	1046	SCMHUDDAMRTEFKQLTITVFAEFARDBPRYALILGDKTRLPA-----TTSODEXOLI	1098
Qy	1019	-DMGDLVDAEYLVPOQGFPCRPDAPGAGWVHHRHSSSTRSGGDLTLGLEBSSEEAR	1077
Db	1099	RKLAPTTDGSBAIKAPDDYLQPKAALGPS-----HRDCT-----DEMP	1137
Qy	1078	-----BSPLAPSEGASDVFDG---DLGMAAGLQSLPETHDPSPIQRISEDPVLPUS	1128
Db	1138	KLNRYCXPDSKNKNSGTGDDERDSSAREGVGNLR-----LDLPV	1176
Qy	1129	ETDGVNAPLTCSPQPEYVNPQDVPPOPSPREGRPLPARPAGATLERAKTLPQKGVK	1188
Db	1177	DEDDYLPN-TCQPGPNNNNNMN-----NPQNNMAVGVAAGYM-----	1214
Qy	1189	DVFAFGAVENBEYL---TPQGAAPQPH-----PPAFSP-AFDN	1225
Db	1215	DLIGVPSVDNREYILNAGTLGVGESPIPTGTIGIPVWGPGGTMEVVKVPMPGSEPTSDH	1274
Qy	1226	LYYWD	1230
Db	1275	EYVND	1279

Search completed: July 22, 2003, 09:10:03  
Job time : 31.0157 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)  
5088.033 Million cell updates/sec

Title: SEQ4-465-479-14

Perfect score: 6842  
Sequence: 1 MELAALCRWGLLLALLPPQA.....TFKGTPTAENPEYLGLDVVPV 1261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6626	96.8	1255	1	ERB2_HUMAN
2	5852	85.5	1257	1	ERB2_RAT
3	5846.5	85.5	1254	1	ERB2_MESAU
4	3116	45.5	1210	1	EGFR_HUMAN
5	3087	45.1	1210	1	EGFR_MOUSE
6	2942.5	43.0	1308	1	ERB4_HUMAN
7	2924	42.7	1308	1	ERB4_RAT
8	2677.5	39.1	1167	1	XMRK_XIPMA
9	2390.5	34.9	1342	1	ERB3_HUMAN
10	2323.5	34.0	1339	1	ERB3_RAT
11	1920	28.1	1426	1	EGFR_DROME
12	1749.5	25.6	634	1	ERBB_ALV
13	1703	24.9	604	1	ERBB_AVIER
14	1630	23.8	540	1	ERBB_AVIEU
15	1571	23.0	703	1	EGFR_CHICK
16	1276	18.6	1323	1	L173_CAEL
17	1142.5	16.7	245	1	ERB2_MOUSE
18	717	10.5	1363	1	ILPR_BRALA
19	702	10.3	1383	1	INSR_RAT
20	701.5	10.3	1372	1	INSR_MOUSE
21	701	10.2	1300	1	INSR_MOUSE
22	701	10.2	1382	1	INSR_HUMAN
23	693	10.1	1297	1	IRK7_HUMAN
24	691	10.1	1477	1	HTK7_HYDAT
25	684.5	10.0	1300	1	IRR_CAVPO
26	680	9.9	1607	1	MIPR_LYMST
27	645	9.4	1367	1	IGIR_HUMAN
28	630	9.2	1390	1	INSR_AEDAE
29	626	9.1	1373	1	IGIR_MOUSE
30	623	9.1	2146	1	INSR_DROME
31	622.5	9.1	1370	1	IGIR_RAT
32	601	8.8	987	1	EPB4_HUMAN
33	597.5	8.7	984	1	EPB1_CHICK

34	597.5	8.7	984	1	EPB1_RAT	P09759	rattus norv
35	591.5	8.6	984	1	EPB1_HUMAN	P54762	homo sapien
36	587.5	8.6	977	1	EPB2_MOUSE	Q03145	mus musculus
37	586	8.6	902	1	EPBB_XENLA	Q01945	mus musculus
38	585.5	8.6	976	1	EPB2_HUMAN	P29317	homo sapien
39	583	8.5	987	1	EPB4_MOUSE	P54761	mus musculus
40	578	8.4	1114	1	RET_HUMAN	P07949	homo sapien
41	573.5	8.4	1053	1	PAK1_CHICK	Q00944	gallus gall
42	572.5	8.4	985	1	EPBA_XENLA	Q01571	xenopus lae
43	569	8.3	1068	1	FAK1_XENLA	Q01736	xenopus lae
44	565.5	8.3	757	1	HTI6_HYDAT	P53356	hydra atten
45	565.5	8.3	986	1	EPBA_XENLA	Q01845	xenopus lae

## ALIGNMENTS

RESULT 1	ERB2_HUMAN	STANDARD;	PRT; 1255 AA.
ID	ERB2_HUMAN		
AC	P04626;		
DT	13-AUG-1987 (Rel. 05, Created)		
DT	13-AUG-1987 (Rel. 05, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell		
DE	surface receptor HER2) (MLN 19).		
GN	ERBB2 OR HER2 OR NGL OR NEU.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86118663; PubMed=3003577;		
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,		
RA	Saito T., Toyoshima K.;		
RT	"Similarity of protein encoded by the human c-erbB-2 gene to		
RT	epidermal growth factor receptor.";		
RL	Nature 319:230-234 (1986).		
RL	(2)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86070181; PubMed=2999974;		
RA	Conseus L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,		
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,		
RA	Frankie U., Levinson A., Ulrich A.,;		
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor		
RT	shares chromosomal location with neu oncogene.";		
RL	Science 230:1132-1139 (1985).		
RN	(3)		
RP	SEQUENCE OF 737-1031 FROM N.A.		
RX	MEDLINE=86016729; PubMed=2995967;		
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.,		
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the		
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a		
RT	human salivary gland adenocarcinoma.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).		
RN	(4)		
RP	VARIANTS VAL-654 AND VAL-655.		
RX	MEDLINE=93194196; PubMed=8095488;		
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;		
RT	"Characterization of a new allele of the human ERBB2 gene by allele-		
RT	specific competition hybridization.";		
RL	Genomics 15:426-429 (1993).		
CC	-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,		
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A		
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-		
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS		
CC	(POTENTIAL).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		

CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY)  
 CC -!- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M11767; AAA35808.1; -  
 CC EMBL: M11761; AAA35808.1; JOINED.  
 CC EMBL: M11762; AAA35808.1; JOINED.  
 CC EMBL: M11763; AAA35808.1; JOINED.  
 CC EMBL: M11764; AAA35808.1; JOINED.  
 CC EMBL: M11765; AAA35808.1; JOINED.  
 CC EMBL: M11766; AAA35808.1; JOINED.  
 CC EMBL: M11730; AAA35808.1; -  
 CC EMBL: M12036; AAA35978.1; -  
 CC EMBL: X03363; CAA27060.1; -  
 CC PIR: A25491; A25491.  
 CC PIR: A24571; A24571.  
 CC HSP: P11362; 1FGK.  
 CC Genew: HGNC:3430; ERBB2.  
 CC MIM: 164870; -  
 CC InterPro: IPR000494; EGFR\_L\_domain.  
 CC InterPro: IPR000719; Euk\_pkinase.  
 CC InterPro: IPR002174; Furin-like.  
 CC InterPro: IPR001245; Tyr\_pkinase.  
 CC InterPro: IPR004019; YLP\_motif.  
 CC Pfam: PF00069; pkinase; 1.  
 CC Pfam: PF00757; Furin-like; 1.  
 CC Pfam: PF01030; Recep\_L\_domain; 2.  
 CC Pfam: PF02757; YLP; 2.  
 CC Prodom: PD000001; Euk\_pkinase; 1.  
 CC SMART: SMO0261; Fuf; 3.  
 CC SMART: SMO0219; Tyrc; 1.  
 CC PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 CC Kinase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Polymorphism.  
 CC KW  
 CC SIGNAL  
 CC CHAIN 1 21 POTENTIAL.  
 CC DOMAIN 22 1255 RECEPTOR\_PROTEIN-TYROSINE\_KINASE\_ERBB-2.  
 CC TRANSMEM 653 652 EXTRACELLULAR (POTENTIAL).  
 CC DOMAIN 676 1255 POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC NP\_BIND 720 987 PROTEIN\_KINASE.  
 CC BINDING 726 753 ATP (BY SIMILARITY).  
 CC ACT\_SITE 753 753 ATP (BY SIMILARITY).  
 CC ACT\_SITE 845 845 BY SIMILARITY.  
 CC DISULFID 195 204 BY SIMILARITY.  
 CC DISULFID 199 212 BY SIMILARITY.  
 CC DISULFID 220 227 BY SIMILARITY.  
 CC DISULFID 224 235 BY SIMILARITY.  
 CC DISULFID 236 244 BY SIMILARITY.  
 CC DISULFID 240 252 BY SIMILARITY.  
 CC DISULFID 255 264 BY SIMILARITY.  
 CC DISULFID 268 295 BY SIMILARITY.  
 CC DISULFID 299 311 BY SIMILARITY.  
 CC DISULFID 315 331 BY SIMILARITY.  
 CC DISULFID 334 338 BY SIMILARITY.  
 CC DISULFID 511 520 BY SIMILARITY.  
 CC DISULFID 515 528 BY SIMILARITY.  
 CC DISULFID 531 540 BY SIMILARITY.  
 CC DISULFID 544 560 BY SIMILARITY.

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	634	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTID=VAR_004077.
FT	VARIANT	655	655	I -> V.
FT	CONFLICT	1170	1170	P -> A (IN REF. 2).
FT	SEQUENCE	1255 AA; 137909 MW; 399DFDA04DF962 CRC64;		
Qy	Query Match	Best Local Similarity	96.8%; Score 6626; DB 1; Length 1255;	
Qy	Matches 1224; Conservative	9; Mismatches 22; Indels 6; Gaps 1;		
Qy	1	MELALCRWGLLLALLPGAASTOVCTGDMKRLPASPETHLDMLRHLYOGCQVQGNL	60	
Db	1	MELALCRWGLLLALLPGAASTOVCTGDMKRLPASPETHLDMLRHLYOGCQVQGNL	60	
Qy	61	ELTYLPTASLFLDIOGVGVLIANOVQVPLQRLIRYRGTLFEDNVALAVLDNG	120	
Db	61	ELTYLPTASLFLDIOGVGVLIANOVQVPLQRLIRYRGTLFEDNVALAVLDNG	120	
Qy	121	DLNNTTPTVGTASPGGLRELQSLTEILKGVLIQRLNQLCYODTILKKDIFHKNOIA	180	
Db	121	DLNNTTPTVGTASPGGLRELQSLTEILKGVLIQRLNQLCYODTILKKDIFHKNOIA	180	
Qy	181	LTLIDTNRSRACHPCSPKCKGRGCESEDCQSLTRVYACAGGACRCKPLPTDCHEOC	240	
Db	181	LTLIDTNRSRACHPCSPKCKGRGCESEDCQSLTRVYACAGGACRCKPLPTDCHEOC	240	
Qy	241	AAGCTGPRKSDCLACLPFHSGICELHCAVLTNTDTESNPNPGRITTFGASCCTACB	300	
Db	241	AAGCTGPRKSDCLACLPFHSGICELHCAVLTNTDTESNPNPGRITTFGASCCTACB	300	
Qy	301	YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLCMOYIKANSKFI	360	
Db	301	YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLCMOYIKANSKFI	360	
Qy	361	ELFPAQCKKIFGSLAPLPESPDGPASNTAPLQPEQLVFTLEBITGYLYISAWPDSL	420	
Db	361	ELFPAQCKKIFGSLAPLPESPDGPASNTAPLQPEQLVFTLEBITGYLYISAWPDSL	420	
Qy	421	DISSVONQVIRGRILHNGAVSLTQIGLISWGLRSLRELSSGLAIHNNHLCVHVY	480	
Db	421	DISSVONQVIRGRILHNGAVSLTQIGLISWGLRSLRELSSGLAIHNNHLCVHVY	480	
Qy	481	PMDQLFRFNNFTVSFMLRVKVSASHLECEVGEGLACHOLCARGHMGCPQCVNCSOF	540	
Db	481	PMDQLFRFNNFTVSFMLRVKVSASHLECEVGEGLACHOLCARGHMGCPQCVNCSOF	540	
Qy	541	LRGQECVEBCRYLQGLPREYVNAHICLPCHPCOPONGSVTCFGBADQCVACAHYKDP	600	
Db	541	LRGQECVEBCRYLQGLPREYVNAHICLPCHPCOPONGSVTCFGBADQCVACAHYKDP	600	
Qy	601	PCVACPSGVKPDLSYMPIMKFPDEGACQPCINCTHSVLDKDGCAEQAPASPLTSI	660	
Db	601	PCVACPSGVKPDLSYMPIMKFPDEGACQPCINCTHSVLDKDGCAEQAPASPLTSI	660	
Qy	661	VSAAVGIILVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI	720	
Db	661	VSAAVGIILVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI	720	
Qy	721	ISAVVGIILVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI	780	
Db	721	ISAVVGIILVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI	780	

```

QY 721 LKTELARKVVLGSGAGCTYKGIWIPDGENVKIPVAIKVIRENTSPANKKEIDEAYVM 780
DB 715 LKTELARKVVLGSGAGCTYKGIWIPDGENVKIPVAIKVIRENTSPANKKEIDEAYVM 774
QY 781 AGVSPVYSRLGICLSTVQLVQLMPLYGCLLDHVRENRRLSODLLNKMOMAKMS 840
DB 775 AGVSPVYSRLGICLSTVQLVQLMPLYGCLLDHVRENRRLSODLLNKMOMAKMS 834
QY 841 YLEEDRVLRHDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPDKMALESI 900
DB 835 YLEEDRVLRHDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPDKMALESI 894
QY 901 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLLEKGERLPPQPICTIDVYMI 960
DB 895 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLLEKGERLPPQPICTIDVYMI 954
QY 961 MVKCMWIDSECRPFRELVSFESRMARDPQRFVYIQLNEDLGPASPLDSTFFRSLLEDDDM 1020
DB 955 MVKCMWIDSECRPFRELVSFESRMARDPQRFVYIQLNEDLGPASPLDSTFFRSLLEDDDM 1014
QY 1021 GDLVDAEEYLVPOQGFCCPDPAAGGVMVHRHRSSTRSGGDLTLGLPSEEEAPRSP 1080
DB 1015 GDLVDAEEYLVPOQGFCCPDPAAGGVMVHRHRSSTRSGGDLTLGLPSEEEAPRSP 1074
QY 1081 LAPSEGAAGSDVFDGDLGMAAKGICLSLPTHDPSPLQRYSEDPTVPLPSETDGYAAPLTCG 1140
DB 1075 LAPSEGAAGSDVFDGDLGMAAKGICLSLPTHDPSPLQRYSEDPTVPLPSETDGYAAPLTCG 1134
QY 1141 POPPYVNOPDVRPOPSPREGPLPAARPAAGTLERAKTSLGKGVVMDVAFGGAENP 1200
DB 1135 POPPYVNOPDVRPOPSPREGPLPAARPAAGTLERAKTSLGKGVVMDVAFGGAENP 1194
QY 1201 EYLTPQGAAPOHPPEAFSPAFDNLVYMDODPPERGAPESTFGTPTAENPEYLGLDVP 1260
DB 1195 EYLTPQGAAPOHPPEAFSPAFDNLVYMDODPPERGAPESTFGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255

```

```

RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofes F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "three dimensional structure of the transmembrane region of the proto-
RL oncogenic and oncogenic forms of the new protein.";
EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP10 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-ch).
CC -----
DR EMBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSSP; P11362; 1EGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Euk_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep-Like; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 1257
FT DOMAIN 22 654
FT TRANSMEM 655 677
FT DOMAIN 678 1257
FT DOMAIN 159 369
FT DOMAIN 473 646
FT DOMAIN 722 989
FT NP_BIND 728 736
FT BINDING 755 755
FT ACT_SITE 847 847
FT DISULFID 196 205
FT DISULFID 200 213
FT DISULFID 221 228
FT DISULFID 225 236
FT DISULFID 237 245
FT DISULFID 241 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 312
FT DISULFID 316 332
FT DISULFID 335 339
FT DISULFID 513 522

```

```

FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT MOD_RSS 632 644 BY SIMILARITY.
FT MOD_RSS 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.5%; Score 5852; DB 1; Length 1257;
Best Local Similarity 85.8%; Pred. No. 1.6e-305;
Matches 1085; Conservative 52; Mismatches 117; Indels 10; Gaps 3;

QY 1 MELALGRMGLLALLPPGAASVCTGTDMLRLPASPTHLMDLRLHLYOGCQVVGNTL 60
D 1 MELALGRMGLLALLPPGIAGTCTGTDMLRLPASPTHLMDLRLHLYOGCQVVGNTL 60
QY 61 ELTYLPNNAISLFIODIOEVGVYLAHNOVRQVPLQRLIRYKGTOLFEDNYALAVLDNG 120
D 61 ELTYLPNNAISLFIODIOEVGVYLAHNOVRQVPLQRLIRYKGTOLFEDNYALAVLDNR 120
QY 121 DPLNNTTPVT-GASPGGLRELQSLTEILKGVLIQNPQLCYODTILMKDIFHKNOL 179
D 121 DPLNNTTPVT-GASPGGLRELQSLTEILKGVLIQNPQLCYODTILMKDIFHKNOL 179
QY 180 ALTLIDNRSRACHPCSPKMGSRKMGSESSDCSLTRTVGAGGACRCKPPLPTDCHEQ 239
D 180 ALTLIDNRSRACHPCSPKMGSRKMGSESSDCSLTRTVGAGGACRCKPPLPTDCHEQ 239
QY 240 CAAGCTGPKSDCLACHFNHSGICEHLCPALVYNTDTFESMPNBEGRYTFGASCVTAC 299
D 240 CAAGCTGPKSDCLACHFNHSGICEHLCPALVYNTDTFESMPNBEGRYTFGASCVTAC 299
QY 241 CAAGCTGPKSDCLACHFNHSGICEHLCPALVYNTDTFESMPNBEGRYTFGASCVTAC 300
D 241 CAAGCTGPKSDCLACHFNHSGICEHLCPALVYNTDTFESMPNBEGRYTFGASCVTAC 300
QY 300 PYNLTSDVSGCTVCPPLHNOEVAEDTORCEKSKCARVCGYLGMOYIKANSKFTGI 359
D 300 PYNLTSDVSGCTVCPPLHNOEVAEDTORCEKSKCARVCGYLGMOYIKANSKFTGI 359
QY 301 PYNLTSTEVSGCTVCPPNNOEVAEDTORCEKSKCARVCGYLGMEHLRGARATSD 360
D 301 PYNLTSTEVSGCTVCPPNNOEVAEDTORCEKSKCARVCGYLGMEHLRGARATSD 360
QY 360 TELEFAGCKIIFGLAPLPESFDGDPASNTAPLQPEQLQVEETLEITGYLISAMPDL 419
D 360 TELEFAGCKIIFGLAPLPESFDGDPASNTAPLQPEQLQVEETLEITGYLISAMPDL 419
QY 361 NVGFQDCKIFGLAPLPESFDGDPASNTAPLQPEQLQVEETLEITGYLISAMPDL 420
D 361 NVGFQDCKIFGLAPLPESFDGDPASNTAPLQPEQLQVEETLEITGYLISAMPDL 420
QY 420 PDLVSFONLQYRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 479
D 420 PDLVSFONLQYRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 479
QY 421 RDLVSFONLQYRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 480
D 421 RDLVSFONLQYRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 480
QY 480 VPMDQLFRFNFTVSFMLRVPKVSASHLE--CYGEGLAGHQLCARGHGCMGPGPCQNC 537
D 480 VPMDQLFRFNFTVSFMLRVPKVSASHLE--CYGEGLAGHQLCARGHGCMGPGPCQNC 537
QY 481 VPMDQLFRFNFTVSFMLRVPKVSASHLE--CYGEGLAGHQLCARGHGCMGPGPCQNC 533
D 481 VPMDQLFRFNFTVSFMLRVPKVSASHLE--CYGEGLAGHQLCARGHGCMGPGPCQNC 533
QY 538 SGLRGECEVEECVLOGLPREYVNAHRLCHPECPONSGVTCFGEADQCAVACAYK 597
D 538 SGLRGECEVEECVLOGLPREYVNAHRLCHPECPONSGVTCFGEADQCAVACAYK 597
QY 534 SHFLRGECEVEECVLMKGLPREYVSDKRLCHPECPONSGVTCFGEADQCAVACAYK 593
D 534 SHFLRGECEVEECVLMKGLPREYVSDKRLCHPECPONSGVTCFGEADQCAVACAYK 593
QY 598 DPPPCVARGSGVVPDLSYPIKFPPEBEGACQPCINCHSCVDLDKGPAPQAPQAPL 657
D 598 DPPPCVARGSGVVPDLSYPIKFPPEBEGACQPCINCHSCVDLDKGPAPQAPQAPL 657
QY 594 DSSSCVARGSGVVPDLSYPIKFPPEBEGACQPCINCHSCVDLDKGPAPQAPQAPL 653
D 594 DSSSCVARGSGVVPDLSYPIKFPPEBEGACQPCINCHSCVDLDKGPAPQAPQAPL 653
QY 658 TSIYSAVAGLILVVLGVFGILIKRROOKIRKTYMRLLOETELVEPIYSGAMPNQAQ 717
D 658 TSIYSAVAGLILVVLGVFGILIKRROOKIRKTYMRLLOETELVEPIYSGAMPNQAQ 717
QY 654 TFIATVAVGILILVVLGVFGILIKRROOKIRKTYMRLLOETELVEPIYSGAMPNQAQ 713
D 654 TFIATVAVGILILVVLGVFGILIKRROOKIRKTYMRLLOETELVEPIYSGAMPNQAQ 713
QY 718 MRIKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVALIKVLRNTPRANKIILDEA 777
D 718 MRIKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVALIKVLRNTPRANKIILDEA 777

```

```

D 714 MRIKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVALIKVLRNTPRANKIILDEA 773
QY 778 YVMAVGSPYVSRLLGICLSTVQLVTLMPYGLLDHVENRGRGSGODLNMCMQIAK 837
D 774 YVMAVGSPYVSRLLGICLSTVQLVTLMPYGLLDHVENRGRGSGODLNMCMQIAK 833
QY 838 GMSVLEDRVLRHRLAANNVYKSPNHVKTIDPGLARLLDIDETEVHADGGVPIKMMAL 897
D 834 GMSVLEDRVLRHRLAANNVYKSPNHVKTIDPGLARLLDIDETEVHADGGVPIKMMAL 893
QY 898 ESIARRRTHSDVMSYGVWELMTFGAKPYDGIIPAREIDLEKGRLLPQPICTIDV 957
D 894 ESIARRRTHSDVMSYGVWELMTFGAKPYDGIIPAREIDLEKGRLLPQPICTIDV 953
QY 958 YMIWKCMMIDSECRPRELVESESRMARPPORFVVIQNEIDLGPASPLDSTFYSLLED 1017
D 954 YMIWKCMMIDSECRPRELVESESRMARPPORFVVIQNEIDLGPASPLDSTFYSLLED 1013
QY 1018 DDMGDIVDAEELYLPQOGFCPPDPAPAGCGVNHHRSSSTRSGGDLTLGLEPSEEAR 1077
D 1014 DDMGDIVDAEELYLPQOGFCPPDPAPAGCGVNHHRSSSTRSGGDLTLGLEPSEEAR 1073
QY 1078 RSPILAPSEGASDVFDGDLGMAKGLQSLPTHPSPLOKRYSEDPVPLPSETDGVAPL 1137
D 1074 RSPILAPSEGASDVFDGDLGMAKGLQSLPTHPSPLOKRYSEDPVPLPSETDGVAPL 1133
QY 1138 TCSPOPEYVNPQVPPSPRECEPLPAARPAATLEBAKTLSPGKNGVNVQVFAFGAV 1197
D 1134 ACSPOPEYVNPQVPPSPRECEPLPAARPAATLEBAKTLSPGKNGVNVQVFAFGAV 1193
QY 1198 ENPEYLPFGGGAAPQPPPAFPAFMDLYYWDOPPERKAPPSFFKKTPTAENEYICL 1257
D 1194 ENPEYLPFGGGAAPQPPPAFPAFMDLYYWDOPPERKAPPSFFKKTPTAENEYICL 1253
QY 1258 DVPV 1261
D 1254 DVPV 1257

RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
AC 06053;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
DN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_Taxid=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishikawa T.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

```



RESIDUES.  
- - - - - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
- - - - -  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
- - - - -  
DR EMBL: D16295; BAA03801.1; -  
DR HSSP: P11362; IFGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_kinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_kinase.  
DR InterPro: IPR004019; YLP\_motif.  
DR Pfam: PF00069; kinase; 1.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR Pfam: PF02757; YLP; 2.  
DR ProDom: PD000001; Euk\_kinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; Tyrc; 1.  
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.  
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KM transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KM proto-oncogene; Disease mutation.  
FT CHAIN 1 21  
FT SIGNAL 1 21  
FT CHAIN 22 1254  
FT DOMAIN 22 652  
FT TRANSMEM 653 675  
FT DOMAIN 676 1254  
FT DOMAIN 158 368  
FT DOMAIN 472 644  
FT DOMAIN 720 987  
FT NP\_BIND 726 734  
FT BINDING 753 753  
FT ACT\_SITE 845 845  
FT DISULFID 195 204  
FT DISULFID 199 212  
FT DISULFID 236 244  
FT DISULFID 240 252  
FT DISULFID 255 264  
FT DISULFID 268 295  
FT DISULFID 299 311  
FT DISULFID 315 331  
FT DISULFID 334 338  
FT DISULFID 511 520  
FT DISULFID 515 528  
FT DISULFID 531 540  
FT DISULFID 544 560  
FT DISULFID 563 576  
FT DISULFID 567 584  
FT DISULFID 587 596  
FT DISULFID 600 623  
FT DISULFID 626 634  
FT DISULFID 630 642  
FT MOD\_RES 1139 1139  
FT MOD\_RES 1247 1247  
FT CARBOHYD 68 68  
FT CARBOHYD 125 125  
FT CARBOHYD 187 187  
FT CARBOHYD 259 259  
FT CARBOHYD 530 530  
FT CARBOHYD 571 571  
FT CARBOHYD 629 629  
FT CARBOHYD 658 658  
FT VARIANT 659 659  
FT VARIANT 659 659  
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.5%; Score 5846.5; DB 1; Length 1254;  
Best Local Similarity 85.3%; Pred. No. 3.2e-305;  
Matches 1080; Conservative 60; Mismatches 109; Indels 17; Gaps 3;  
QY 1 MELALCRMGILLALLIPGAASTOYCTGDMKRLPASETHLMDRLHYOGCOVQGNL 60  
DB 1 MELAMCGGILLALLIPGASGTQCTGDMKRLPASETHLMDRLHYOGCOVQGNL 60  
QY 61 ELTYVPTNASLSFLDIOGVGVLIANHOVQVRLRIYRGTOLEFDNVALATLNDG 120  
DB 61 ELTYVPANATLSFLDIOGVGVLIASHOVHVPRLRIYRGTOLEFDNVALATLNDR 120  
QY 121 DPLNNTPTVGTASPGGLRELQRLSTEILKGVLIORNPQLCYODITLMDIPIHKNQOLA 180  
DB 121 DPLDNTTATGRTPEGLRELQRLSTEILKGVLIORNPQLCYODITLMDIPIHKNQOLA 180  
QY 181 LTLIDTNRSRACHPCSPKCKGSRCWGESSEDCQSLTRIVCAGGACRCKPLPTDCHEQC 240  
DB 181 PVDIDTNRSRACHPCAPACKNDKNCWASPEDCOTLGTIAPRAVPAARARLPDCCHEQC 240  
QY 241 AAGCTGPRKSDCLACHFHNSGICELHCAALVTYNTDTESMPNPEGRTTFGASCTYACP 300  
DB 241 AAGCTGPRKSDCLACHFHNSGICELHCAALVTYNTDTESMPNPEGRTTFGASCTYTCP 300  
QY 301 VYVLTSDVGSCTLVCPRLHNOEYTAEDGTORCKGSPCARVCYGLGMOYIKANSKFIGIT 360  
DB 301 VYVLTSDVGSCTLVCPRLHNOEYTAEDGTORCKGSPCARVCYGLGMOYIKANSKFIGIT 360  
QY 361 ELEFAGCKKIFGSLAFLEPSPDGPASNTAPLOPELQVFEETLEITGYLYISAMPDLSL 420  
DB 361 IEFAGCKKIFGSLAFLEPSPDGPASNTAPLOPELQVFEETLEITGYLYISAMPDLSL 420  
QY 421 DLSVQNTQVYGRILHNGAYSLTQGLISWGLRSLRELGSGLAIHHNTHLCFVHTV 480  
DB 421 DLSVQNTQVYGRILHNGAYSLTQGLISWGLRSLRELGSGLAIHHNTHLCFVHTV 480  
QY 481 PMDOLFRRNFVFWLRVPKVSASHL-----EECGEGLACHQACARHCWGPGTQCV 535  
DB 481 PMDOLF-----RNPHQALHSGNPSEEGCLQDPACYPCLCAHCHWGPGTQCV 535  
QY 536 NCSOPLRQGECEVEGRVLOGLPREYVNAHCLPCHPECOPONGSVTCFPEADQCVACAH 595  
DB 536 NCSHRLRQGECEVEGRVLOGLPREYVNAHCLPCHPECOPONGSVTCFPEADQCVACAH 595  
QY 596 YKDPFCVACRCPGKVPDLSYPIWKFPDEGACQPCPINCTHSCVYDLDDKGPABQRAS 655  
DB 596 YKDSPFVACRCPGKVPDLSYPIWKFPDEGACQPCPINCTHSCVYDLDDKGPABQRAS 655  
QY 656 PLTISVAVGILLVVLGVVFGILLIKRQOKIRKTYMRALLQETELVPLTPSGAMPNO 715  
DB 656 PATSIITATVGLLFLVIGVAVGILLIKRQOKIRKTYMRALLQETELVPLTPSGAMPNO 715  
QY 716 AQMRILKTELKRVKVLGSGAFGYKGIWIPGENVKIPVAKVLRKENTSPKANEILD 775  
DB 716 AQMRILKTELKRVKVLGSGAFGYKGIWIPGENVKIPVAKVLRKENTSPKANEILD 775  
QY 776 EAYVAGVSPVYSRLGICLTSTVOLVTLMPYCLLDHVENRGRGLSODLNNCMQI 835  
DB 776 EAYVAGVSPVYSRLGICLTSTVOLVTLMPYCLLDHVENRGRGLSODLNNCMQI 835  
QY 836 AKGMSYLEDVRLVHVDLAARNVLYSPNNVKTIDGRLARLDIDETEHADGKVPYIKM 895  
DB 836 AKGMSYLEDVRLVHVDLAARNVLYSPNNVKTIDGRLARLDIDETEHADGKVPYIKM 895  
QY 896 ALESILRRRTFQSDVMSYGVTVMEIPLFGAKPYGCIIPAREIPDLLEKGERLPQPICTI 955  
DB 896 ALESILRRRTFQSDVMSYGVTVMEIPLFGAKPYGCIIPAREIPDLLEKGERLPQPICTI 955  
QY 956 DVTMIMVCMWIDSECRPRFELVSEFSRMARDPQRFVITQNEDLGSPASPLDSTFYRSL 1015  
DB 956 DVTMIMVCMWIDSECRPRFELVSEFSRMARDPQRFVITQNEDLGSPASPLDSTFYRSL 1015

QY 1016 EDDMDGLVDAEELVLPQGGFCPPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEE 1075  
DB 1010 EDDMDGLVDAEELVLPQGGFCPPDPAAGAGSTAHRRHSSSTRSGGDLTLGLEPSEEE 1069  
QY 1076 APRSLAPSGAGSDVFDGLGMGAAGKLOSLPHDPSPLQRISEDPVLPSEDTGYVA 1135  
DB 1070 PPRSLAPSGAGSDVFEGLGMGATKGPQSIIPRDLSPQRYSEDPVLPSEDTGYVA 1129  
QY 1136 PLTCSPOPEXVNPDPVPPSPREGPLPAPAPAGATLERAKTLSPGKGVVMDVFAFGG 1195  
DB 1130 PLACSPQPELVNPPVPPPLPPEGLPVPVPAGATLERPKTLSPKNGVMDVFPFGG 1189  
QY 1196 AVENPEYLTPQGAAPQPPPPAFAFDNLVYWDOPPERGAPSPSTFKGTPPAENPEYL 1255  
DB 1190 AVENPEYLVRGSGASQPP-PPALCPAFDNLVYWDQPPSRGSPPTFECTPAENPEYL 1248  
QY 1256 GLDVPV 1261  
DB 1249 GLDVPV 1254

RESULT 4  
EGFR\_HUMAN STANDARD; PRT; 1210 AA.  
ID EGFR\_HUMAN Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;  
AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;  
AC 000688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
protein-cytosolic kinase ErbB-1).  
GN EGFR OR ERBB1.  
OS Homo sapiens (Human).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=84219729; PubMed=6328312;  
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,  
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
RT expression of the amplified gene in A431 epidermoid carcinoma cells";  
RT Nature 309:418-425(1984).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Placenta;  
RX MEDLINE=95382957; PubMed=7654368;  
RA Ilekis J.V., Stark B.C., Scoccia B.;  
RT "Possible role of variant RNA transcripts in the regulation of  
RT epidermal growth factor receptor expression in human placenta";  
RL Mol. Reprod. Dev. 41:149-156(1995).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Placenta;  
RX MEDLINE=97078686; PubMed=8918811;  
RA Reiter J.L., Maibhe N.J.;  
RT "A 1.8 kb alternative transcript from the human epidermal growth  
RT factor receptor gene encodes a truncated form of the receptor";  
RL Nucleic Acids Res. 24:4050-4056(1996).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Placenta;  
RX MEDLINE=97256547; PubMed=9103388;  
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;  
RT "Expression of a truncated epidermal growth factor receptor-like  
RT protein (TEGFR) in ovarian cancer";  
RL Gynecol. Oncol. 65:36-41(1997).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
RC TISSUE=Placenta;  
RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
RA Schehl Sinclair C., Pearrell R.S., Green P.J., Yee D., Lampland A.L.,  
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
RA Maibhe N.J.;  
RT "Comparative genomic sequence analysis and isolation of human and  
RT mouse alternative EGFR transcripts encoding truncated receptor  
RT isoforms";  
RL Genomics 71:1-20(2001).  
RN [6]  
RP SEQUENCE OF 575-687 FROM N.A.  
RA Reiter J.L., Thredgill D.W., Danielsen A.J., Schehl C.M.,  
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
RA Maibhe N.J.;  
RT "Human and mouse alternative EGFR transcripts encoding only the  
RT extracellular domain of the receptor";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE OF 713-924 FROM N.A.  
RX MEDLINE=84196372; PubMed=6326261;  
RA Lin C.R., Chen W.S., Krulger W., Stolarsky L.S., Weber W.,  
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
RT "Expression cloning of human EGF receptor complementary DNA: gene  
RT amplification and three related messenger RNA products in A431  
RT cells";  
RL Science 224:843-848(1984).  
RN [8]  
RP SEQUENCE OF 150-962 FROM N.A.  
RX MEDLINE=84245835; PubMed=6330563;  
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
RA Roe B.A., Merlino G.T., Pastan I.;  
RT "Human epidermal growth factor receptor cDNA is homologous to a  
RT variety of RNAs overproduced in A431 carcinoma cells";  
RL Nature 309:806-810(1984).  
RN [9]  
RP SEQUENCE OF 1028-1210 FROM N.A.  
RX MEDLINE=85046483; PubMed=6093780;  
RA Simmen F.A., Cope M.L., Schulz T.Z., Wright D.A., Carpenter G.,  
RA O'Malley B.W.;  
RT "Isolation of an evolutionarily conserved epidermal growth factor  
RT receptor cDNA from human A431 carcinoma cells";  
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
RN [10]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=88217333; PubMed=3329716;  
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
RA Waterfield M.D.;  
RT "The human EGF receptor gene: structure of the 110 kb locus and  
RT identification of sequences regulating its transcription";  
RL Oncogene Res. 1:375-396(1987).  
RN [11]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=91107677; PubMed=1988448;  
RA Haley J.D., Waterfield M.D.;  
RT "Contributory effects of de novo transcription and premature  
RT transcript termination in the regulation of human epidermal growth  
RT factor receptor proto-oncogene RNA synthesis";  
RL J. Biol. Chem. 266:1746-1753(1991).  
RN [12]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=85270438; PubMed=2991899;  
RA Ishii S., Xu Y.H., Straton R.H., Roe B.A., Merlino G.T., Pastan I.;  
RT "Characterization and sequence of the promoter region of the human  
RT epidermal growth factor receptor gene";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
RN [13]  
RP SEQUENCE OF 540.  
RA Kohda D.;  
RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
RN [14]  
RP RECEPTOR ACTIVITY.  
RX MEDLINE=84191554; PubMed=6325948;  
RA Mroczkowski B., Mosig G., Cohen S.;  
RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and supercoiled DNA.";  
 RL Nature 309:270-273(1984).  
 RN (15)  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Marcolis B.L., Lax I., Kries R., Dombalagian M., Hongger A.M.,  
 RT Howk R., Givol D., Ullrich A., Schlessinger J.,  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor.";  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN (16)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.,  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts.";  
 RL Growth Factors 13:121-132(1996).  
 RN (17)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.,  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor.";  
 RL J. Biochem. 127:65-72(2000).  
 RN (18)  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Otake M., Inagaki F., Lax I., Schlessinger J., Kohda D.,  
 RT "Disulfide bond structure of human epidermal growth factor receptor.";  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN (19)  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.,  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens.";  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -I- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -I- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -I- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -I- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -I- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X00588; CAA25240.1; -  
 CC EMBL; U95089; AAB53063.1; -  
 CC EMBL; U48722; AAC50802.1; -

DR EMBL; U48723; AAC50804.1; -  
 DR EMBL; U48724; AAC50796.1; -  
 DR EMBL; U48725; AAC50797.1; -  
 DR EMBL; U48726; AAC50798.1; -

Query Match 45.5%; Score 3116; DB 1; Length 1210;  
 Best Local Similarity 49.5%; Pred. No. 2,7e-159;  
 Matches 630; Conservative 177; Mismatches 350; Indels 116; Gaps 25;

QY 11 LILALTPGAA--STVCGTGTDMKLLPASPETHLDMLRHLVGGCVGVGNLELTYPNTN 68  
 DB LILALTPGAA--STVCGTGTDMKLLPASPETHLDMLRHLVGGCVGVGNLELTYPNTN 68  
 QY 14 LLAALCPASRALEEKVKVCGSTSNKLTQGLGFEDHFFSLQGFNNFNCVGLNLETTYQRN 73  
 DB LLAALCPASRALEEKVKVCGSTSNKLTQGLGFEDHFFSLQGFNNFNCVGLNLETTYQRN 73  
 QY 69 ASLSFLQDIQEVGVYLLAHNQVYPLQRLIRVGTQLPEDVYALAVLNDGPNLNTTP 128  
 DB ASLSFLQDIQEVGVYLLAHNQVYPLQRLIRVGTQLPEDVYALAVLNDGPNLNTTP 128  
 QY 74 YDLSFKTIQEVAVGYLLANTYERLPLENLQIIRGMYYENSYALAVLSNYD----- 126  
 DB YDLSFKTIQEVAVGYLLANTYERLPLENLQIIRGMYYENSYALAVLSNYD----- 126  
 QY 129 VTGASPGIARELQRLSLTEILKGVLIQRPOLCYODTIWKQI FHKNNQALTLIDTNR 188  
 DB VTGASPGIARELQRLSLTEILKGVLIQRPOLCYODTIWKQI FHKNNQALTLIDTNR 188  
 QY 127 ---ANKTGKELPMRNLQETILHGAIVRFSNNPALCNVESIQWRDIVSDFLSNMSDFQNH 183  
 DB ---ANKTGKELPMRNLQETILHGAIVRFSNNPALCNVESIQWRDIVSDFLSNMSDFQNH 183  
 QY 189 SRACHSPCMCKGSRGCMGSESSDQSLTRTVAGCGA-RCKGPLPDDCCECAAGCTGP 247  
 DB SRACHSPCMCKGSRGCMGSESSDQSLTRTVAGCGA-RCKGPLPDDCCECAAGCTGP 247  
 QY 184 LGSQCKQDPSCPNGSCWGAEBECCQKLTIKCAQCSGRGKSPSDCCNQCAGCTGP 243  
 DB LGSQCKQDPSCPNGSCWGAEBECCQKLTIKCAQCSGRGKSPSDCCNQCAGCTGP 243  
 QY 248 KHSQDLACLFHNSGICELCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNYLSTD 307  
 DB KHSQDLACLFHNSGICELCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNYLSTD 307  
 QY 244 RSDCLVCRKFRDEACXKCTPLMLYNPTTYQMDVNPBCKYSFGATCYVCKCRNRYVTD 303  
 DB RSDCLVCRKFRDEACXKCTPLMLYNPTTYQMDVNPBCKYSFGATCYVCKCRNRYVTD 303  
 QY 308 VSGCTLVCPRLHNOETVAEDGTORCEKSKRPARVCYGLQMOYIKANSKFIGITELE-FAG 366  
 DB VSGCTLVCPRLHNOETVAEDGTORCEKSKRPARVCYGLQMOYIKANSKFIGITELE-FAG 366  
 QY 304 HSCVACACADSYEM-EEDGVKCKKCEGRCRVKNGCIGEFK-DLSINATNIKAFKN 361  
 DB HSCVACACADSYEM-EEDGVKCKKCEGRCRVKNGCIGEFK-DLSINATNIKAFKN 361  
 QY 367 CKKIFSLFLPSPFGCDASNPALQEPQLOVFELEETGTGLYSIAMDLSPLDSVFG 426  
 DB CKKIFSLFLPSPFGCDASNPALQEPQLOVFELEETGTGLYSIAMDLSPLDSVFG 426  
 QY 362 CTISGDLHILPAPFGDSFTHTPLPDLDELDTLKVKETGFLLIQAMENNTDLAFB 421  
 DB CTISGDLHILPAPFGDSFTHTPLPDLDELDTLKVKETGFLLIQAMENNTDLAFB 421  
 QY 427 NLQVIRGRILHNGAVSLTQGLGISWIGRLSRLRELSGLALIHNNTHLCFVHTVPDQLF 486  
 DB NLQVIRGRILHNGAVSLTQGLGISWIGRLSRLRELSGLALIHNNTHLCFVHTVPDQLF 486  
 QY 422 NLEIIRGRIRKQGFSLAVASLINTSLGRSLKEISDGVYIISGNRLCYANTINMKLFL 481  
 DB NLEIIRGRIRKQGFSLAVASLINTSLGRSLKEISDGVYIISGNRLCYANTINMKLFL 481  
 QY 487 RNNFTVSFWLAPVKSASHLE-ECVGEGLACHQLCARGHGMGPGPTOCVNCQSQPLRGQE 545  
 DB RNNFTVSFWLAPVKSASHLE-ECVGEGLACHQLCARGHGMGPGPTOCVNCQSQPLRGQE 545  
 QY 482 GTSG-----OKTKIISNRGENSCATQGVCHALCSPBCQWBPBPDVCSGNVSGRE 534  
 DB GTSG-----OKTKIISNRGENSCATQGVCHALCSPBCQWBPBPDVCSGNVSGRE 534  
 QY 546 CYBECKVLOGLRREYVNAHCLPCHECQPPQNSVTCFSPBADQVCAAHYKPPPCVAR 605  
 DB CYBECKVLOGLRREYVNAHCLPCHECQPPQNSVTCFSPBADQVCAAHYKPPPCVAR 605  
 QY 535 CVDKCNLEGEEREFENSECTIQCHECULPQAMNITCTGRGPNCITQCAHYIDGPHCVKT 594  
 DB CVDKCNLEGEEREFENSECTIQCHECULPQAMNITCTGRGPNCITQCAHYIDGPHCVKT 594  
 QY 606 CPGVKPDLSYMPIMKFPDEGACQPCPINCCHSCVDLDDKGPAPGASPLTSISAVY 665  
 DB CPGVKPDLSYMPIMKFPDEGACQPCPINCCHSCVDLDDKGPAPGASPLTSISAVY 665  
 QY 595 CPAGVWGENNTL-VMKYADAGVCHLCHNCTYGTGPGLEGCTNGPKIP--SIATGMV 651  
 DB CPAGVWGENNTL-VMKYADAGVCHLCHNCTYGTGPGLEGCTNGPKIP--SIATGMV 651  
 QY 666 G---LLVVVLGVFGLIKRRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAKRIILK 722  
 DB G---LLVVVLGVFGLIKRRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAKRIILK 722  
 QY 652 GALLLLLVVALGIG--LFMRBRHIVRKTLRLLOERELVEPLTPSGAMPNALLRIILK 708  
 DB GALLLLLVVALGIG--LFMRBRHIVRKTLRLLOERELVEPLTPSGAMPNALLRIILK 708  
 QY 723 FTELKRVKLVGSGAGTGVKGIWIPDGEVVKI PVAIKVLRNTSPPAKNEILDEAYVMAG 782  
 DB FTELKRVKLVGSGAGTGVKGIWIPDGEVVKI PVAIKVLRNTSPPAKNEILDEAYVMAG 782  
 QY 709 ETEFKKIKVLGSGAGTGVKGIWIPDGEVVKI PVAIKVLRNTSPPAKNEILDEAYVMAS 768  
 DB ETEFKKIKVLGSGAGTGVKGIWIPDGEVVKI PVAIKVLRNTSPPAKNEILDEAYVMAS 768  
 QY 783 VGSFVYSRLGICLFTSTQVLVQLMPYGLDLHVRENRRGLSODLLNMCQIAKMSVY 842  
 DB VGSFVYSRLGICLFTSTQVLVQLMPYGLDLHVRENRRGLSODLLNMCQIAKMSVY 842  
 QY 769 VDNPHVCRLLGICLSTSTQVLVQLMPYGLDLHVRENRRGLSODLLNMCQIAKMSVY 828  
 DB VDNPHVCRLLGICLSTSTQVLVQLMPYGLDLHVRENRRGLSODLLNMCQIAKMSVY 828  
 QY 843 EDVRLVHNDLAAARNVLYVSPNFKITDFGLARLDIDETRYHADGGKVPKIMMALESILR 902  
 DB EDVRLVHNDLAAARNVLYVSPNFKITDFGLARLDIDETRYHADGGKVPKIMMALESILR 902  
 QY 829 EDRLRVHNDLAAARNVLYVSPNFKITDFGLARLDIDETRYHADGGKVPKIMMALESILR 888  
 DB EDRLRVHNDLAAARNVLYVSPNFKITDFGLARLDIDETRYHADGGKVPKIMMALESILR 888  
 QY 903 RRFTHOSDVMSGYVWELMTFGARPYDGIIPAREIPDLLEKKEKRLPQPICTIDVYMIW 962  
 DB RRFTHOSDVMSGYVWELMTFGARPYDGIIPAREIPDLLEKKEKRLPQPICTIDVYMIW 962  
 QY 889 RIYTHOSDVMSGYVWELMTFGSKPYDGIIPASEISISILEKGERLPQPICTIDVYMIW 948  
 DB RIYTHOSDVMSGYVWELMTFGSKPYDGIIPASEISISILEKGERLPQPICTIDVYMIW 948

Qy	963	ICMMIDSCRPRLVSEFSRMADPORFVYIO-NEDLGASPLDSIFYSLLEDMDG	1021
Db	949	KCMWIDASRKFRLLIIEFSKMDRFORVYIIGDEKMHLPSPIDSNFTALNDEMD	1008
Qy	1022	DLVDAEYLYVPOQGFCCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEA	1081
Db	1009	DVDAEYLYVPOQGF-----SSPSTRSRL	1034
Qy	1082	APSEGAGDVFDGDLGGMGAKGLSLPTNHPSPLORYSEDPVLPSEST--DGVA	1139
Db	1035	LSLSLSTSN--NSTYACIDRNGLOSCPKEDSFLORYSSDPGLALTEISIDDTFL	1087
Qy	1140	SPOREYVYNQPDVRFOPSPFREGPLPAARPAGATLERAKTLSPGKGVKDVPA	1199
Db	1088	PVEEYIYNQ-SVPRKRPAGSVNPPVHNPILNP--APSRDPHYOD--PSTAGVN	1135
Qy	1200	PEYV-TPQSGAAROPHRPAFSPADNLUYUDQ-----DP-----PERGAP	1242
Db	1136	PEYLVNQ-----PTCVNSTSTDSRAHMAQKSHQISLDNPVYQDDPFKKA	1166
Qy	1243	FKGTPTAENPEYL 1255	
Db	1187	FKGS-TAENAEYL 1198	

ID	EGFR_MOUSE	STANDARD;	PRT;	1210 AA.
AC	001279;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112).			
GN	EGFR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C; TISSUE=Liver;			
RX	MEDLINE=93026370; PubMed=1408137;			
RA	Avivi A., Skorecki K., Yayon A., Givol D.;			
RT	"Promoter region of the murine fibroblast growth factor receptor 2			
RL	(b6k/KGFR) gene.";			
RL	Oncogene 7:1957-1962(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C, and CD-1; TISSUE=Liver, and Decidua;			
RX	MEDLINE=93126380; PubMed=7678348;			
RA	Paria B.C., Das S.K., Andrews G.K., Dey S.K.;			
RT	"Expression of the epidermal growth factor receptor gene is regulated			
RL	in mouse blastocysts during delayed implantation.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C; TISSUE=Liver;			
RA	Hibbs M.L.;			
RL	Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=B6/C3; TISSUE=Liver;			
RX	MEDLINE=94170986; PubMed=8125255;			
RA	Luetsteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,			
RT	Jenkins N.A., Lee D.C.;			
RL	"The mouse waved-2 phenotype results from a point mutation in the EGF			
RL	receptor tyrosine kinase."			
RL	Genes Dev. 8:399-413(1994).			
RN	[5]			
RP	SEQUENCE OF 1-714 FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=91232866; PubMed=2030916;			
RA	Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;			

```

RT      "Comparison of EGF receptor sequences as a guide to study the ligand
RT      binding site."
RT      Oncogene 6:673-676 (1991) .
RN      [6]
RN      RP      SEQUENCE OF 969-1117 FROM N.A.
RC      STRAIN=CH:
RA      Eisinger D.P., Serrero G.:
RL      Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC      AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC      VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY) .
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC      dimerization, internalization of the EGF-receptor complex,
CC      induction of the tyrosine kinase activity, stimulation of cell DNA
CC      synthesis, and cell proliferation.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch) .

```

CC	EMBL	X78987	CAA55587.1	-	-
DR	EMBL	U03425	AAA17899.1	-	-
DR	EMBL	X59698	CAA42219.1	-	-
DR	EMBL	L06864	AAA53029.1	-	-
DR	EMBL	Z12608	CAA78249.1	-	-
DR	HSSP	P11362	1FGK	-	-
DR	MCD	MGI:95294	Egfr	-	-
DR	InterPro	IPR000494	EGFR_L_domain	-	-
DR	InterPro	IPR000719	Euk_Pkinase	-	-
DR	InterPro	IPR002174	Furin-like	-	-
DR	InterPro	IPR001245	Tyr_pkinase	-	-
DR	Pfam	PF00069	Pkinase.1	-	-
DR	Pfam	PF00757	Furin-like.1	-	-
DR	Pfam	PF01030	Recep_L_domain.2	-	-
DR	ProDom	PD000001	Euk_Pkinase.1	-	-
DR	SMART	SM00261	FU; 3.	-	-
DR	SMART	SM00219	TYRK; 1.	-	-
DR	PROSITE	PS00107	PROTEIN_KINASE_ATP; 1.	-	-
DR	PROSITE	PS00109	PROTEIN_KINASE_TYR; 1.	-	-
DR	PROSITE	PS50011	PROTEIN_KINASE_DOM; 1.	-	-
KM	Transmembrane		Glycoprotein; Receptor; Signal; Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.	-	-
KM	Transmembrane		Glycoprotein; Receptor; Signal; Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.	-	-
FT	SIGNAL	1	24	POTENTIAL.	-
FT	CHAIN	25	1210	EPIDERMAL GROWTH FACTOR RECEPTOR.	-
FT	DOMAIN	25	647	EXTRACELLULAR (POTENTIAL).	-
FT	TRANSMEM	648	670	POTENTIAL.	-
FT	DOMAIN	671	1210	CYTOPLASMIC (POTENTIAL).	-
FT	REPEAT	75	300	APPROXIMATE.	-
FT	REPEAT	390	600	APPROXIMATE.	-
FT	DOMAIN	1028	1071	SER-RICH.	-
FT	DOMAIN	714	981	PROTEIN KINASE.	-
FT	NP BIND	720	728	ATP (BY SIMILARITY).	-
FT	BINDING	747	747	ATP (BY SIMILARITY).	-
FT	ACT_SITE	839	839	BY SIMILARITY.	-
FT	DISULFID	190	199	BY SIMILARITY.	-
FT	DISULFID	194	207	BY SIMILARITY.	-
FT	DISULFID	215	223	BY SIMILARITY.	-
FT	DISULFID	219	231	BY SIMILARITY.	-
FT	DISULFID	232	240	BY SIMILARITY.	-
FT	DISULFID	236	248	BY SIMILARITY.	-
FT	DISULFID	251	260	BY SIMILARITY.	-
FT	DISULFID	264	291	BY SIMILARITY.	-
FT	DISULFID	295	307	BY SIMILARITY.	-
FT	DISULFID	311	326	BY SIMILARITY.	-
FT	DISULFID	311	326	BY SIMILARITY.	-
FT	DISULFID	329	333	BY SIMILARITY.	-

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 991 991 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF22F5 CRC64;

```

Query Match 45.1%; Score 3087; DB 1; Length 1210;  
 Best Local Similarity 49.4%; Pred. No. 9.6e-158;  
 Matches 632; Conservative 170; Mismatches 358; Indels 120; Gaps 27;

```

11 LLLALLPAGAA--STVCTGTDMLRLPASEETLDMRLHYQCCQVVOGNETLTYPTN 68
14 LLLALCAAGALEBKVCQGSNRLTQLGTFEDHFLSLQRYNNCEVVLGLLETTYQORN 73
69 ASISFLDIOEVQGYVLIANNQVQVPLQRLRIVRGTQLFEDNALVLDNGPLANTTP 128
74 YDLSFLKTIQEVQGYVLIANNQVQVPLQRLRIVRGTQLFEDNALVLDNGPLANTTP 124
129 VTGASPGGLRELQRLSLTEILKGVLLIQRPOLCYDTIIMKDI----FKNNQALATLI 184
125 -YGTNRGLRELPMRNLOEILIGAVRSPNPILCNMDTIQWRDIYQVWFMNSMDL--- 180
185 DTRRSAPCHPCSPMKGSRCKWGESSEDCQSLTRVCGGCA-RCKGPLPDDCHEQCAAG 243
181 -QSPSSCPKCDSPNGSCWGGEEENCKLTKLICQOCSHRCGRSPSDCHNQCAAG 239
244 CTGPKISDCLACLHFNHSGICELHCPALVTYNTDTEFSPMNPREGVTFGASCVTACPNY 303
240 CTGPRSDCLVCKQFQDEATCTCPPLMLYNTPTTYMDVNPBGKTSFGATCVAKKCPRY 299
304 LSTDVSCSLVCPRLHNOEVTABDGTORCEKSKRCARVCGYLGWQYKANSKFGITELE 363
300 VVTDHSCSVACGCPDYEV-EEQGIKCKCKDGCRCRVCGIGIGERK-DTSLINATNIK 357
364 -PAGCKKIFGSLAFLESFPGDASNTAPLOPELOVFTLEBITGYLYISAMPDLPDL 422
358 HFKYCTAISGDHLILPVAFKGDSTFTRPPLDPRELEILTKVKEITGFLILQAMPDWTDL 417
423 SVFQNTQVIRGRILHNGAVSLTLOGIGISWLGRLSRLRELSGLALIHNNHLCFVHTVPM 482
418 HAFENLEIIRKIRKQKQGFSLAVGANTISLGRSLKEISDGVITISGNBNLCYANTINM 477
483 DQLFNNFTVSWLWPKVSAHLE-ECYEGELACHOLCARGHCMGPGPTOCVNSQFL 541
478 KKLFGTPN-----OKTKIMNRAREKDKAVNHVNCPLCSSGCGPREDVCVQSNVS 530
542 RQGECEBEKRVLOGLRREVYNNAHCLPCHPECQPNQSGVTCFGBADQCVACAHYDDPP 601

```

```

Db 531 RGRCEVCENCLIEBEPREFVENSECIOCHPECLPQAMNITCTGKRPDNCIOCAHYIDPH 590
602 CVACPSGVKDLSTYPMWKPEDEGACQPCPINCTHS CVLDKDGCPAEGRASPSTIV 661
591 CVKTCPPGIMENNTL-VMKYADANNCHLCHACCTGACGPGLOGCEVWPSGPKIPISA 649
662 SAVGILLVVVLGVVFGI-LIKRRQOKIRKTYMRLLOETELVEFLTPSGAMPNOQRI 720
650 TGIYGLGLFIV-VALIGLFMRRRHIVKRTLRLLQERLVEFLTPSGEAPNAHLRI 708
721 LKETELRKVYKLSGAGFTGYKGIWIDGENVKIPALIKVLRENTSPRANKELIDEAYM 780
709 LKETEFKIKVIGSGAGFTGYKGIWIDGENVKIPALIKVLRENTSPRANKELIDEAYM 768
781 AGVSPVSRILGICLTSTVOLYQMLPYGCLLDHVENRRLSODLLNMCQIAKMS 840
769 ASVNPVPCRLGICLTSTVOLYQMLPYGCLLDHVENRRLSODLLNMCQIAKMN 828
841 YLEBDRVLVHRDLAARNVLYKSPNHVKITDFGLARLIDITETRYADGGKVPILKMALESI 900
829 YLEBRRLVHNDLAARNVLYKTPQHVKITDFGLARLIDITETRYADGGKVPILKMALESI 888
901 LRRRFTQSDVMSYGVTVWELMTFGAARPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
889 LHRITYHQSDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPICTIDVYMI 948
961 MVKCMIMDSRCRPRFRRLVSEFSMARDPQRFVYIQ-NEDGPRSPLDSTYRSLIEDD 1019
949 MVKCMIMDASRPFRRLVSEFSMARDPQRFVYIQDERHRLSPDTSNFEYRALMDEED 1008
1020 MGDVLDAEYLVLPQGFPCFDPAPGAGMYHHRSSSTRSGGDLTLGLPSEEARNS 1079
1009 MEDVYVDELVLIPQGF-----NSST-----SRT 1034
1080 PLAPSEGAGSDVEDGDLGMAKAGLSLPTHPSPLOYSEDPVLPSEF-DGYVAPL 1137
1035 PLTSLISATSN---NSTVACINRNGSCRVKEDAFLORYSSDPGAVTEIDIDAFL--- 1087
1138 TCSPPQRYVNOVDVRRPSPRESREPLPAARPAAGTLERAKTSLSGKGVKVDVAFGCAV 1197
1088 ---PVPEYVNO-SVPKRPAGSVQNPVYHNPFLP-----APGRDLHYON--PHSNVAV 1133
1198 ENPEYL-TPOGGAAPQHPAPFAPFADNYUWQ-----DP-----PERGAP 1240
1134 GNPBYLNTAQ-----PTCLSSGPNPALMTIQKSHQMSLDNPYQODFFPKETPN 1184
1241 STFKGTPTAENPEYGLDVP 1260
1185 GIFKG-PTAENAEYLRVAP 1203

```

RESULT 6  
 ERBA HUMAN STANDARD; PRT; 1308 AA.  
 ID ERBA HUMAN  
 AC 015303;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase etbB-4 precursor (BC 2.7.1.112)  
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).  
 GN ERBA OR HER4.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX SEQUENCE FROM N.A. (ISOFORM JM-A).  
 RP TISSUE=Breast carcinoma;  
 RX MEDLINE=93189574; PubMed=8383326;  
 RA Pionman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,  
 Foy L., Neubauer M.G., Shoyab M.;  
 "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

Query Match	Similarity	Score	294.2.5	DB 1	Length	1308
Best Local	Similarity	45.0%	Pred. No. 5.5e-150			
Matches	609	Conservative	181	Mismatches	383	Indels 179; Gaps 29;
QY	9	WGLLLALLPEPAA-----STOVCTGTDMKLRPLPASPEFHLDMLRLHYGCCOVGONLELTY	64			
DB	8	WWWVSLVAVGATVPDSDGSVCACTENKLSLSLDEQYBALRKYENCCEVVMGNLEITS	67			
QY	65	LPTVASLSPLODIOEVGVYLIANOVROYPLQGLRIYRGTOLEPDENVYALAVLDNGDPLN	124			
DB	68	IEHNRDLSFLRSVLEVGYVLVNLNQRYPLPLENLRIRITGKLIEDYALALFLNTRKDG	127			
QY	125	NTTDPVTGASPEGIREILOLRSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNOLATLI	184			
DB	128	NF-----GLQELKILKULTEILNGVYVDQKFLCYADTIHMQIVANPMPNSLTLV	178			
QY	185	DYNRSRACHPCSPMCKSSRCMGBSESDCCSLTRVYVACGC-AACKGFLPDDCCHEQCAG	243			
DB	179	STNSSSGCGRHKHSCSTG-RCMGPEFNHCQTLLTRVVCACQDGDGRYGGYVADCCRCBAG	237			
QY	244	CTGPKHSDCLACLFHNSGILCELIHCPLALVTYNTDTFESMNPREGRYFFGASCVYACPYNT	303			
DB	238	CSGKXDIDFCACMKNFNDSGACVYQCQPTFLVYNPTTFLBENHFNPAKTYIYGAFVYKCKCHNF	297			
QY	304	LSTDVGSCTLVCPILANOEVTAEDGTQRCCKSCSKPCARVYGLGMOYIKANSKFIIGTELE	363			
DB	298	V-VDSSSCVACPSKSKIEV-EENGIXKCKECTDIPCACGIGTGSILMSAQTVDSNSNDK	355			
QY	364	PACGKTKFGSLAFIPESFDGDPASNTPLDPEOLOVETLEETIGYLIYSAMPDSDLDS	453			

```

Db 356 FINCTKINGNLFLVTGTHGDPYNAIEAIDPEKLNVEFTVREITGELINIQSHPNNTPDS 415
Qy 424 VFONLOVIRGRILHNGAVSLTQGLISWLGRLSRELSGGLALIHNTHLGCVHTVPM 483
Db 416 VFSNLVTIGGRVLYSGLSLILKQGITSLQFOSLKEISAGNIYIDNSNLCTYHNT 475
Qy 484 QLFRRNNTVSWLVRPKVKSASHLEECVGEGLACHOLCARGHCWGPFPQCVNCSQPLRG 543
Db 476 TLFFSTIN-----ORIVIRDRNRKAENCTAEGWVCHNLCSDDCWGPGPQCLSCRRFSRG 529
Qy 544 QECVECEVLQGLPREVYVNAHCLPCHPCOP-ONGSVYCFGEADQCAVHYXDP 602
Db 530 RLCHSCNLVDEFEFEFENGSTICVBCDPQCEHDEGLTCHGFPUNCTKSHFKDGPNC 589
Qy 603 VARCPGVKPDLSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKGC-----PAE 651
Db 590 VEKCPDGLQGANFS--IFKYADPRDECHCHNCCTGCGCPNSHDCITYYPMTGSHLPRH 647
Qy 652 QASPLTSIVASV-GLLVVVLGVVFGILIRROOKIRKYMRLLOETELVEPLTPSG 710
Db 648 AR-TPL--IAGVIGLFLVIVGLFAVYVRKRSIK-KKRALRRL-ETELVEPLTPSG 702
Qy 711 AMPNAGMRLKETELRKVKVLSGAFGVYKGIWPDGENYKIPAIIVLAENSPKAN 770
Db 703 TAPNQAQRLKETELRKVKVLSGAFGVYKGIWPDGENYKIPAIILNETTGPKAN 762
Qy 771 KEILDEAVYMGVSGPVYSRLIGICTSTVQLVTOLMPYGCLLDHYRNGRLSGDILLN 830
Db 763 VEFMEDEALIMASMDPHLVRLGLVCLSPITQLVTOLMPHGLCLEYHEHKNDISGOLLN 822
Qy 831 WCMQIAKMSYIEDVRLVHRDLAARNVLYKSPNNHYKITDGLARLLIDETEVHADGRV 890
Db 823 WCVQIAKMSYIEERLVRDLAARNVLYKSPNNHYKITDGLARLLIDETEVHADGRV 882
Qy 891 PIKMALESLIRRFTHOSDWSYGVTWELMTFPAKPYDGI-PAREIPLLEKGERLPP 950
Db 883 PIKMALESLIRRFTHOSDWSYGVTWELMTFPAKPYDGI-PAREIPLLEKGERLPP 942
Qy 951 PICTIDVYIMVKKCMWIDSECRPRFELVSEFSRMAPDQRFVNIQED-LGPASPLDST 1009
Db 943 PICTIDVYIMVKKCMWIDSECRPRFELVSEFSRMAPDQRFVNIQED-LGPASPLDST 1002
Qy 1010 FYRSLIEDDDMDGLVDAEBEYLPQCGFCPPDPAAGCVNTHRRSSSTRSGGDLTGL 1069
Db 1003 FFONLDEEDLEDMDAEEYLP-QAFNIPRP-----IYTSRAIDSNRS-----EIGH 1050
Qy 1070 EPSEEARPS-----PLAP-SEGASGVVDGDLGMA 1100
Db 1051 SPPEAYTMSGNQFYRDGGAEGCVSVYPAPTSTIPBAVAQCATLEIFDSCCNGT 1110
Qy 1101 AKGLQSLPTHPSPLOQRYSEDPTEPLPS-----ETDGYVALTCSQPEVYNQDVP 1153
Db 1111 LRKPAVPHVQEDSSQRYASDPTVAPERSPRGELDEBQYMPMDKPROEVLNVE--- 1167
Qy 1154 QPSPREPLPAARAGATLEBAKTLSPEKGVKQVAFGAVENPEYLPQGAAPOR 1213
Db 1168 -----ENPFVSR- -KNGDLQ-----ALNPEYHNASNG----- 1194
Qy 1214 HPPPA-----FSPAPNLYWODDPPERGA- 1238
Db 1195 -PPKAEDYVNEPLYLNTFANTLGAELYKNNIISMPEKAKAFDNPDMNSLPPRSTL 1253
Qy 1239 -PPSTFKGTP-----AENPEYL 1255
Db 1254 QHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

```

```

RESULT 7
ERB4 RAT STANDARD; PRT; 1308 AA.
ID ERB4 RAT
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulin promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN (2)
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN (3)
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
RN (4)
RP FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
RN 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
RN NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
RN NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
RN CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
RN tyrosine phosphate.
RN SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
RN RECEPTORS (POTENTIAL).
RN SUBCELLULAR LOCATION: Type I membrane protein.
RN TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
RN NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
RN OF EXPRESSION IN THE ADULT BRAIN WHERE IT IS MAINLY FOUND IN THE
RN RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
RN HEART.
RN PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
RN RESIDUES (BY SIMILARITY).
RN SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN -----
RN EMBL; AF041838; AAD08899.1; -
RN EMBL; U52531; AAC53051.1; -
RN HSSP; P11362; 1FGK.
RN InterPro; IPR000494; EGFR_L_domain.
RN InterPro; IPR000719; Euk_pkinase.
RN InterPro; IPR002174; Furin-like.
RN InterPro; IPR001245; Tyr_pkinase.
RN InterPro; IPR004019; YLP_motif.
RN Pfam; PF00757; Furin-like; 1.
RN Pfam; PF00069; Kinase; 1.
RN Pfam; PF01030; Recep_L_domain; 2.
RN Pfam; PF02757; YLP; 2.
RN PRINTS; PR00109; TYRKINASE.
RN ProDom; PD000001; Euk_pkinase; 1.

```



DR SMART; SM00261; FU; 4.  
DR SMART; SM00219; TYKIC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; signal;  
KW transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 25  
FT CHAIN 26 1308  
FT DOMAIN 26 651  
FT TRANSMEM 652 675  
FT DOMAIN 676 1308  
FT DOMAIN 186 334  
FT DOMAIN 496 633  
FT DOMAIN 718 985  
FT NP\_BIND 724 732  
FT BINDING 751 751  
FT ACT\_SITE 843 843  
FT DISULFID 189 197  
FT DISULFID 193 205  
FT DISULFID 213 221  
FT DISULFID 217 229  
FT DISULFID 230 238  
FT DISULFID 234 246  
FT DISULFID 249 258  
FT DISULFID 262 289  
FT DISULFID 293 304  
FT DISULFID 308 323  
FT DISULFID 326 330  
FT DISULFID 503 512  
FT DISULFID 507 520  
FT DISULFID 523 532  
FT DISULFID 536 552  
FT DISULFID 555 569  
FT DISULFID 559 577  
FT DISULFID 580 589  
FT DISULFID 593 614  
FT DISULFID 617 625  
FT DISULFID 621 633  
FT MOD\_RES 1162 1162  
FT MOD\_RES 1188 1188  
FT MOD\_RES 1258 1258  
FT MOD\_RES 1284 1284  
FT CARBOHYD 138 138  
FT CARBOHYD 174 174  
FT CARBOHYD 253 253  
FT CARBOHYD 358 358  
FT CARBOHYD 410 410  
FT CARBOHYD 473 473  
FT CARBOHYD 495 495  
FT CARBOHYD 548 548  
FT CARBOHYD 576 576  
FT CARBOHYD 620 620  
FT CONFLICT 1062 1062  
FT CONFLICT 1080 1082  
SQ SEQUENCE 1308 AA; 146957 MW; D944BB096A08B41 CRC64;  
Query Match 42.7%; Score 2924; DB 1; Length 1308;  
Best Local Similarity 44.9%; Pred. No. 5,4e-149;  
Matches 607; Conservative 188; Mismatches 392; Indels 166; Gaps 29;  
QY 1 MEIA-ALCRWGLL--ALPPGAASOVCTGTMTKRLRPASPTLDMRLRLHLYOGGOWO 57  
DB 1 MKLATGLWVGSLVLAARTVQPSASQSVACGTENKLSLSLDEQVRLARKYENKEVVM 60  
QY 58 GULELTYPFNASLSLQDIQEVQVYLIAHNOVROVPLQRLRIIVGTQLFEDNVALVL 117  
DB 61 GULETISIEHNRLSLRSIREVTGYVALNDFYLPLENLRIIRGKLYEDRYALALF 120  
QY 118 DNGDDPLNNTTPTVTSAPGGLRELOSLRLTEILKGGVLIOQNPOLCVQDITIMQDIHKAN 177  
DB 121 LMYRKDGNF-----GLQELGLKLTLEILNGGVYVDONKFLCYADITIMQDIVRNPM 171

QY 178 QULATLIDTNRSPACHPCSPMGKSRCKGESSSEDCOSLTRVCAGGC-ARCKGRLPTDCC 236  
DB 172 PSMTTLVSTIGSSCCGRCHSKCTG-RCMGTENHCOTLITTVCAECQDGCYGPVYSDCC 230  
QY 237 HEQCAAGCTGPKRSDCLAFHNSGICELHCPALVTYNTDTEMSMPREGRYTFGASCV 296  
DB 231 HRCACGCGSCPKDTCFACNRFNDSGACVTCQCFOTFYVNTPTQLHNNFAKYTGAFV 290  
QY 297 TACPYYLSTDVSCSLVCPYLHNOETAEQTCRCEKSPCARVCYGLCMQYIKANSKF 356  
DB 291 KCPNHFV-VDSSSCVRACPSRMEV-EENGIMCKRCPTDIPCACGIGTGLMSQTV 348  
QY 357 IGITLFPAGCKKIFGSLAFLPESFPDQDASNTAPLOPELOVETLEITGYLYISAMP 416  
DB 349 DSNINDKFINGCTINGNLFLVTGCHGDPNALDAIDPEKLVNFRYREITGFLNTQTP 408  
QY 417 DSLPDLVSFONLOVIRGRILHNGAYSLLQGLGISWGLSLRSLGSLALIHNTNLCF 476  
DB 409 PMNTDPSVFNLTIGGRVLYSGLSLLIKQGGITSLQFSLKEISAGNIYITDNSNLCY 468  
QY 477 VHTVPDOLFRRNFTVSRFLVFKVSASHLECEVGBGLACHOLCARHGCWGPPTCCVN 536  
DB 469 YHTINNTTLTFSTVN-----QRIVIRDNRAENCTAEGMVCNHLCSNDGCGPQDCLS 522  
QY 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPEQCP-ONGSVTCFGEADQVACAH 595  
DB 523 CRFSSKCTIESCNLYDGEFRFENGSLICEVDSQCEKMDGLTCHGQPDCTKCSH 582  
QY 596 YKDPPEVCARCPGVPDLSYMPIMKFPDEGACORPCINCTHSCVDLDDKGC----- 648  
DB 583 FKQGPCVCEKCPVLQGANF--IFKADDRCHPCHRCPTGCGNPTSHDCTLYPMTG 640  
QY 649 ----PDEQASPLTYSVAVV-GILVVVLGVVFGILIKRQOKIRKTYMRLLQETELV 703  
DB 641 HSTLPQPHAR-TPL-IAAGVIGLFLVLMALTFVAVVRRKSIK-KKRALRFL-ETELV 695  
QY 704 EPLTPSGAMPNOQMILKETELRKVYVGSAGAGTYKGIWIPDGENVYKIPVAKYURE 763  
DB 696 EPLTPSGTAPNOQMILKETELRKVYVGSAGAGTYKGIWIPDGENVYKIPVAKYURE 755  
QY 764 NTSRKANKELIDEAAYVAGVSPVSRLLGICLTSTVOLVTOIMPYGLLDHYENRGRL 823  
DB 756 TTPRKANVEMDBALIMASVDHPLVRLIGVCLSPITQVTOIMPYGLLDHYENRGRL 815  
QY 824 GSDDLNCMQIANKGSIYLEDVLYNRDLAARVLYKSPNHYKITDFGLARLLDIDTEY 883  
DB 816 GSQLLNMCVQIAKGMVYLEERLVNRDLAARVLYKSPNHYKITDFGLARLLDIDTEY 875  
QY 884 HADGKVPIKMALESILRRPFTHSDVMSYGTVMELMFGAKKPYDGPAREIPDLEK 943  
DB 876 NADGGMPIKMALESILRRPFTHSDVMSYGTVMELMFGAKKPYDGPAREIPDLEK 935  
QY 944 GERLPQRPICITDVIMYKCMWIDSECRPRFRELSEFSRMAADPQRFVIONED-LGP 1002  
DB 936 GERLPQRPICITDVIMYKCMWIDSECRPRFRELSEFSRMAADPQRFVIONED-LGP 995  
QY 1003 ASPLDSTFRLSLDEDDMDGLVDAAEYLVPQGFPCPD----- 1041  
DB 996 PSFNDKSKFONLDEDEDDMDAAEYLVP-QAFNIPRPYIYSTRIDSNRSEIGHSP 1054  
QY 1042 --APGAGGVNHHNRSSSTRSGGDLTLGLRSEEEEARSPRLAPSGAGSDVVDGLGMC 1099  
DB 1055 AITPMGSGQFVYDGGFATQGG--NPMRYATATSTIPERVA--QGATLMEFDDSCNG 1109  
QY 1100 AAKGLQSLPTHPSPLOQYSEDPVLPDPS-----ETDGYVPLTCSPOPEYVNOQDVR 1152  
DB 1110 TLKRPVPHVVOEDSSQYRSADPTVAPERNRPALEDESGYPMMDKPKQVFLNPE-- 1167  
QY 1153 PQRPSRREGRLPAARPAQATLEBAKTLSPGKGVKQDVAFPGAVENPEYLTQGGAAAPQ 1212  
DB 1168 -----ENPFPVSR-----KNGDLO-----ALDNPENHSASSG----- 1194  
QY 1213 PHPPRA-----FSRPAFDNLVYWDPPPERGA 1238



```

Db      1195 --PPKADEYNEBLYLNTFTNALGNAEYKNSLSLSEBEKAKAFDNDNDYWNHSLPRST 1252
Oy      1239 --PSTFGKPT-----AENPEYL 1255
Db      1253 LQHPDYUOEYSTKYFYKONGRIPLVAENPEYL 1285

RESULT 8
XMRK_XIPMA STANDARD: PRT: 1167 AA:
ID      XMRK_XIPMA
AC      P13388;
DT      01-JAN-1990 (Rel. 13, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN      XMRK OR TY.
OS      Xiphophorus maculatus (Southern platyfish).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC      Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC      Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX      NCBI_TaxID=8083;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90015140; PubMed=2797166;
RA      Witbrodt J., Adam D., Malitschek B., Maueiler W., Raulf F.,
RA      Telling A., Robertson S.M., Scharl M.;
RT      "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT      inducing Tu locus in Xiphophorus.";
RL      Nature 341:415-421 (1989).
RN      [2]
RP      REVISION TO 515.
RA      Scharl M.;
RL      Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC      or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; X16891; CAA34770.2; -
DR      PIR; S06142; S06142.
DR      HSSP; P13362; IRGK.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_Pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR002290; Ser Thr_Pkinase.
DR      InterPro; IPR001245; Tyr_Pkinase.
DR      Pfam; PF00069; Pkinase; 1.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PD000001; Euk_Pkinase; 1.
DR      SMART; SMO0261; FU; 5.
DR      SMART; SMO0230; S_TKC; 1.
DR      SMART; SMO0219; TYKC; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KM      Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KM      Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT      SIGNAL 1 25
FT      CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

```

```

FT      DOMAIN 26 642 KINASE.
FT      TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
FT      DOMAIN 666 1167 POTENTIAL.
FT      DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
FT      BINDING 710 977 PROTEIN KINASE.
FT      BINDING 716 724 ATP (BY SIMILARITY).
FT      ACT_SITE 743 743 ATP (BY SIMILARITY).
FT      ACT_SITE 835 835 BY SIMILARITY.
FT      DISULFID 195 204 BY SIMILARITY.
FT      DISULFID 199 212 BY SIMILARITY.
FT      DISULFID 220 228 BY SIMILARITY.
FT      DISULFID 224 236 BY SIMILARITY.
FT      DISULFID 237 245 BY SIMILARITY.
FT      DISULFID 241 253 BY SIMILARITY.
FT      DISULFID 256 265 BY SIMILARITY.
FT      DISULFID 269 296 BY SIMILARITY.
FT      DISULFID 300 311 BY SIMILARITY.
FT      DISULFID 315 330 BY SIMILARITY.
FT      DISULFID 333 337 BY SIMILARITY.
FT      DISULFID 504 513 BY SIMILARITY.
FT      DISULFID 508 521 BY SIMILARITY.
FT      DISULFID 524 533 BY SIMILARITY.
FT      DISULFID 537 553 BY SIMILARITY.
FT      DISULFID 556 569 BY SIMILARITY.
FT      DISULFID 560 577 BY SIMILARITY.
FT      DISULFID 593 615 BY SIMILARITY.
FT      DISULFID 618 626 BY SIMILARITY.
FT      CARBOHYD 622 634 BY SIMILARITY.
FT      CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 39.1%; Score 2677.5; DB 1; Length 1167;
Best Local Similarity 44.9%; Pred. No. 7.1e-136;
Matches 572; Conservative 167; Mismatches 390; Indels 145; Gaps 29;

Oy      4 AALRGWGLLALLPRGAAS-----QVCTGTDMKLRPLSPERHLDMLHLVGGCOVQGN 59
Db      8 AALLQ--LLLVLSISRCSTDPDRVCQGTSNQMTM---LDNRHYLMKMKMYSGCNVAVLEN 62
Oy      60 LEITYLPTNASTLSEFLODIOEVGYVLIHNOVQVPLQRLRIVRGTOLFEDNYALAVLDN 119
Db      63 LEITYQENQDLSFLQSIQEVGYVLIHNEVSTIPLVNLRIKQNLVEGNFTLLVNSN 122
Oy      120 GDPINNTPTVGTASPGGIRELQLSLTLEILKGVLIOHNPOLCYODTLLMDIFHKNNQL 179
Db      123 YQK--NPSSP--DYQVQGLKQQLSVLTETILSGVYVSHNPPLCNVETINMMDIVDXTSNP 179
Oy      180 ALTLIDTRSRACHPSCMCKGSKCGSSSEDCQSLRTTVACAGG--AACGKPLPDDCCE 238
Db      180 TNLILPFAFERQCKCHGCVNGSCWAPGHCOKFTLLCAEOCNRRCRGKRPIDCCNE 239
Oy      239 QCAAGCTGPKSKDCLAFHNSGICELCPALVTYNTDTESMNPBGRYTFGASCYTA 298
Db      240 HCAGGCTGPRATDCLACDFDDGTCKOTCPKPKIYDIVSHQVVDNPIKTFGAACVXE 299
Oy      299 CPYNYLSTDVSGCTLVCEPLHNOEYTABDGTORCEKSKPCARVCYGLGMOYIKANSKEIG 358
Db      300 CPSNYVYTE--GACVRSASAGMLEVD--ENGRKSKPCDVCPCVKDGGIGSL--SNTIAVN 356
Oy      359 ITLE--EPAGCKKIKGSLAFBESPDGPASTAPLQEPQLOVFEITLITGLYLIASAMP 417
Db      357 STNIRSFSCNCKKINGDIILNNSFEFGDPHYKIGTMDPEHLNLTIVKEITGLYLVIMWPE 416
Oy      418 SLPLDSVFQNLQVIRGRILHNQAYS--LTLLQGLISWLGRLSRLTSGSLALIHNNTHLCF 476

```

```

Db      417  NMTSLVFQNLLEIINGRTTFSGFSFVVVQVRLHQLMGLRSLKEYSAGNVILLKNTLQLRLY 476
      477  VHTVWMDLFRFNNNTVSFWLVRPKVASHLEECVGEGLACGLQCARHGCMARPTQCVN 536
      477  ANTIIMWRLLFSESDSIEYDART-----ENOTCNNESSDDCMWPGPTMCVS 523
      537  CSQFLRGOECVEECVAVLQGLPREVYNARHCLPCHPECCOPQNSVTCFGEADQCVACAHY 596
      524  CLHVDGRGCVAASCHLQGEPRERADVDGRVCVCHQEBLVQVTSILCTYGFGRANCKSAHF 583
      597  KDPFCVAVRCSGVKPRDLSYPMWKPRPEAGACORPCINCTHSCVLDLDCGPABORASP 656
      584  ODGPQCIIRCPHIGLIGDGDTL-IMKYADMGGCCCHQCTGGCGSPGLSGCRGD-IYSH 641
      657  LTSIVSAVGLLVVAVLGVVFGILLKRPQOKIRKXTMRLLQETELVEPLTSSGAMPNA 716
      642  SSLAAGLVSGLLITVALLVLLVLLRRRRIR-KRKTIRLLQEKLEVEPLTSPGAPNOA 700
      717  QMRILKETELRKVKVLSGAGFQTVYKGIWIPDGENVKIPVAIKVIRENTSPANKETLDE 776
      701  FLRIKETEFKKDRVLGSGAGCTVYKGLMNPGENIRIPVAIKVIRENTSPANKETLDE 760
      777  AYWAGVSPYVSRLLGICLTSTVOLVQLMPYGLLDHVRNRRGLSQDLLNMCQOJA 836
      761  AYWASVDPHPYCRLLGICLTSAVOLVQLMPYGLLDHVRNRRGLSQDLLNMCQOJA 820
      837  KGMVLEDEVRLLVHRLAARNVLYKSPNHYKITDPLGLAHLDDIDEHYHADGKRVIKMA 896
      821  KGMVLEDEVRLLVHRLAARNVLYKSPNHYKITDPLGLAHLDDIDEHYHADGKRVIKMA 880
      897  LESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLKGERLLPQPICTID 956
      881  LESILQWYTHQSDVMSYGVTVWELMTFGSKRYDDIPAKELASVLENGERLPQPICTIE 940
      957  VYMWKCMWIDSECRPPRELVSFSSPMARDPQFVVIQNEIDGLPASPLDSTFYRSLLE 1016
      941  VYMIILKCMWIDPSSRPRFRELVEFSSQARDPSRYLYQG--NLPSIDRLRFLSRLLS 997
      1017  DDDMGDLVDAEEYLVPOGFCFOPRAPAGAGVHHRRHSSSTRSGGGLTLGLEFSEEA 1076
      998  SDD--DVVDADDEYLYPKRI-----NRQSS----- 1020
      1077  PRSLPASEGAGSDVFDGLMGAKAGLQSLPTHPDSPLOQRYSEDPV-PIPSETDGYVA 1135
      1021  --EFCIPETGH-----PYRNSITLKNISPTQNALEKLDLGH-- 1056
      1136  PLTCSPOPEYVNOGVRPOP-----PSPRE--GFLP-AARPAATLERAKTILSPG 1182
      1057  -----EYVNOGSETSSRLSDIYNPNYEDLTDGMGPVSLSQEARTFSRPEYLTNTN 1108
      1183  KNGVVKDVFAGAVENPEYLLTPOGGAAPQHPHPAPFAPDNLVYMWODDPERGAPST 1242
      1109  QNSL--PLVSSGSNDPDY--QAG-----YQAAF-----LPQTGALTGN 1143
      1243  FKQPTAENPEYLG 1256
      1144  GMFLPAENLEYLG 1157

```

## RESULT 9

```

ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Iseling W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Pedraza G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.,
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erb3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC - ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM. EXIST DUE TO ALTERNATIVE SPLICING.
CC - TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P65 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC - DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1RGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Kinase.
DR InterPro; IPR002174; Puri-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

```

KM Alternative splicing.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1342  
 FT DOMAIN 20 643  
 FT TRAMEM 644 664  
 FT DOMAIN 665 1342  
 FT DOMAIN 709 966  
 FT NP\_BIND 715 723  
 FT BINDING 742 742  
 FT ACT\_SITE 834 834  
 FT DISULFID 186 194  
 FT DISULFID 190 202  
 FT DISULFID 210 218  
 FT DISULFID 214 226  
 FT DISULFID 227 235  
 FT DISULFID 231 243  
 FT DISULFID 246 255  
 FT DISULFID 259 286  
 FT DISULFID 290 301  
 FT DISULFID 305 320  
 FT DISULFID 323 327  
 FT DISULFID 500 509  
 FT DISULFID 504 517  
 FT DISULFID 520 529  
 FT DISULFID 533 549  
 FT DISULFID 552 565  
 FT DISULFID 556 573  
 FT DISULFID 576 585  
 FT DISULFID 589 610  
 FT DISULFID 613 621  
 FT DISULFID 617 629  
 FT CARBOHYD 126 126  
 FT CARBOHYD 250 250  
 FT CARBOHYD 353 353  
 FT CARBOHYD 408 408  
 FT CARBOHYD 414 414  
 FT CARBOHYD 437 437  
 FT CARBOHYD 469 469  
 FT CARBOHYD 522 522  
 FT CARBOHYD 566 566  
 FT CARBOHYD 616 616  
 FT VARSPLIC 141 183  
 FT  
 FT VARSPLIC 184 1342  
 FT CONFLICT 560 560  
 FT CONFLICT 1064 1064  
 SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;  
 Query Match 34.9%; Score 2390.5; DB 1; Length 1342;  
 Best Local Similarity 40.0%; Pred. No. 1.8e-120;  
 Matches 530; Conservative 196; Mismatches 448; Indels 151; Gaps 36;  
 QY 10 GLTLLALPRGAA--STOYCTGDMTLRLPASPETHLMDLRLHYOGCQVYQCNLELTLYPT 67  
 DB 11 GLTSLARGSEVNSQAVCPGLNGLSVTGAENQYQTLTKYERCEVVMGMLTLYTGH 70  
 QY 68 NASTSLFDIOEVOGYVLIANQVQVPLQRLRIYRGTLPEFNVALLVLDNGDPLNNTT 127  
 DB 71 NADSLFLQWIREVGYVLIANNEFSTLPLRLRVKRGVYVDGKALIVM-----LNNNT 125  
 QY 128 PVTGASRGGLRELOLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLATLLIDTN 187  
 DB 126 ----NSSHALRLQLTQLTEILSGVYIEKNDKLCMDTIDPRDIVRD--AETVVKD 178  
 QY 188 RSRACHPCSPMKCSRCWGESSEDOSLTRTVACAGC-ARKCGPLPTCCHEQCAAGCTG 246  
 DB 179 NGRSCPPCHEVCKG-RCWGESSEDOYLTKTICAQCNCHCFGPRPNQCCCHDECAAGCGSG 237  
 QY 247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPREGRYTFGASCYTACPYNYST 306  
 DB 238 PPDTCFCACRFHNSGACVPRCPQPLVYNNKLTJFQLEPRPHTKYQYGGVASCPRNFV-V 296

QY 307 DVGSCITLVCPRLHNOETAEEDGTORCEKSKPCARVCYGLQMYIKANSKE--IGITELE- 363  
 DB 297 DQSSCRACPRDMEVD-KNGLMKCPGGLCPKACBEGT-----SGSRQYDSSNIDG 350  
 QY 364 PACCKKIPGSLAFPLPSFPQDPASTAPLQPEOLQVETLEBITGYLYISAMPDLS 423  
 DB 351 FVNCSTKLNDLFLTGLNGDPWHKIPALDPKLANFRTRVETITGYLNIQSWPHMNF 410  
 QY 424 VQONLQVIGRIHNGAVS-LTLQGLGIMLGRLSELQSLALHNNHNLGCVHTVPW 482  
 DB 411 VFSNLTTIGRSLYNGFSLILMKNLNVSLGRSLKEISAGRIYSANQLCYHSLNW 470  
 QY 483 DQLFRENNFTVFWLRVP---KVSASH---LEECVEGLACIOLCARGHCGKPTQCN 536  
 DB 471 TVY-----LRGPTERRLDITGNRPDRDCAVEKVCDDLSCSGCGWGPQGLS 519  
 QY 537 CSQFLRGQCEVBCRYLOGLPREYVNAHRLCPHCEPCQPNQSVTCFGEPAQVACAHY 596  
 DB 520 CRYNSRGVCTVHCNPLNGEPREFAEACFCSCHPCEQMEGTATCNGSGSDTCAQAHF 579  
 QY 597 KQPPFCVACPSGVPRDLSYMPIMKRPDEGACQPCPICTHSCVLDLDDKGCFAEORA-- 654  
 DB 580 RQGPFCVSSCPHGVLG--AKGPIYKYPDVQNECRPCHECTQCKGPELODCLQGLTVLI 637  
 QY 655 --SPLTSIVSAVGLILVVVLGVFGILIKRROOKIR-KYTMRLLOETELVPLPSSGA 711  
 DB 638 GKTHLTMLTAVAG--LVYIPMLGTFLYMGRRIQNKRAKRYLERGESIEPLDS-E 694  
 QY 712 MNQAOMLIKETELRKVKVLGSGAFYVYKGIWIPDGENVKIPVAIKYLRNTSPKANK 771  
 DB 695 KANKYLIARIFKETELRKLKVLGSGVGVYHKGWVIEGESIKIPVCIKYIEDSGQSFO 754  
 QY 772 ELDEAYVWAGSVYVSRLGICLTSTVQLYTQMLPYGCLLDHRENGRLQSODLNM 831  
 DB 755 ATDHNMLAGSLDHAIYVLLGLCPSSQLQVLYQLPLGSLLDHVNQHGALCPOLLNM 814  
 QY 832 CMQIAKGSYLEDVRLVHRDLAARVLYVSPHNVKITDGLRLDIDETEVHADGKVP 891  
 DB 815 GVQIAKGYVLEHGMVHRNLAAARVLLKSPQOVYADGVADLPPDDKQLLYSEAKTP 874  
 QY 892 IKMMALESLRRRFTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951  
 DB 875 IKMMALESLHFKYTHQSDVMSYGVYVWELMTFGAEPYGLRLAEVRLDLLEKGERLAQ 914  
 QY 952 ICTIVVMIMVCMNIDSECRFPRELYSESRMARDPQRFVYIQNEIDGLPA--SPLDS 1008  
 DB 935 ICTIDVVMVCMNIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPH 993  
 QY 1009 TFRSLLEDDMDGLVDABEYLVVPOGGFCPRPARGAGMNHRRSSSTRSGGDLTIG 1068  
 DB 994 GLTNKLEFEVLEBELDLDLEABED-----NLATTLTGSLSLP 1034  
 QY 1069 LEP-SEEBAPRPLAPSEGAGSDVFDGLGMGAAGLQSLPTHN-PSPLQRYSEDPVPL 1126  
 DB 1035 VQTLNRPAGSGSLSPSSGY-MPMNQNLGSSCSGSAVSSGERCRPVSILH-----FM 1087  
 QY 1127 P-----SETQGYVA-----PLTCSPOPE---YVNPQVVRPQPSPRE 1160  
 DB 1088 PRGLASSSSRGVYSGAEILOEKVSMCRSRSRSPRPRDSAYNSQHSILTVTPLS 1147  
 QY 1161 GP-----LPAARPAQTLERAKTLP-GKQGV-----KDVAFAGAVENPEY 1202  
 DB 1148 PRGLEBEVNGVMDPTLTKGTPSRBGLTSLVGLSVLGTBEED-----EEVY 1199  
 QY 1203 LTPQGAAPQHPAPAFSPARDNLTYND-----ODPRPGAPPTSTFKGTPRAE 1250  
 DB 1200 MARRRRHSP-RHPRPSSLBELEGVRYMDVGSLSASLSTQSCPLHPRPIMTACTTPE 1258  
 QY 1251 NPEYL 1255  
 DB 1259 DYETM 1263

RESULT 10  
ERB3\_RAT STANDARD; PRT; 1339 AA.  
ID\_062759; 062955; 37, Created)  
AC\_15-DEC-1998 (Rel. 41, Last sequence update)  
DT\_15-JUN-2002 (Rel. 41, Last annotation update)  
DE\_15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)  
(c-erbB3).  
GN ERB3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=liver;  
RX MEDLINE=96096535; Pubmed=8522190;  
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
RT "Cloning of the rat ErbB3 cDNA and characterization of the  
RT recombinant protein";  
RL Gene 165:279-284(1995).  
RN [2]  
RP REVISIONS TO 85; 513 AND 565.  
RA Hellyer N.J., Koland J.G.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 922-1097 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
RX MEDLINE=97184212; Pubmed=9010624;  
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neuroregulating and their putative receptors, Erb2 and  
RT Erb3, is induced during Wallerian degeneration.";  
RL J. Neurosci. 17:1642-1659(1997).  
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.  
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC - SUBCELLULAR LOCATION: Type I membrane protein.  
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
CC PHOSPHATIDYLINOSITOL 3-KINASE.  
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC -----  
DR EMBL; U29339; AAC28498.2; -;  
DR EMBL; U52530; AAC53050.1; -;  
DR HSSP; P11362; 1FGK.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR PRINTS; PP000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00239; TyKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 19  
FT CHAIN 20 1339  
FT DOMAIN 20 643  
FT TRANSMEM 644 662  
FT DOMAIN 663 1339  
FT DOMAIN 183 259  
FT DOMAIN 707 964  
FT NP BIND 713 721  
FT BINDING 740 740  
FT ACT\_SITE 832 832  
FT DISULFID 186 194  
FT DISULFID 190 202  
FT DISULFID 210 218  
FT DISULFID 214 226  
FT DISULFID 227 235  
FT DISULFID 231 243  
FT DISULFID 246 255  
FT DISULFID 259 286  
FT DISULFID 290 301  
FT DISULFID 305 320  
FT DISULFID 323 327  
FT DISULFID 500 509  
FT DISULFID 504 517  
FT DISULFID 520 529  
FT DISULFID 533 549  
FT DISULFID 556 573  
FT DISULFID 576 585  
FT DISULFID 589 610  
FT DISULFID 621 629  
FT DISULFID 617 629  
FT CARBOHYD 126 126  
FT CARBOHYD 250 250  
FT CARBOHYD 353 353  
FT CARBOHYD 408 408  
FT CARBOHYD 414 414  
FT CARBOHYD 437 437  
FT CARBOHYD 469 469  
FT CARBOHYD 522 522  
FT CARBOHYD 566 566  
FT CARBOHYD 616 616  
FT CONFLICT 1028 1028  
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CRC64;  
  
Query Match 34.0%; Score 2323.5; DB 1; Length 1339;  
Best Local Similarity 40.4%; Pred. No. 6,9e-117;  
Matches 523; Conservative 171; Mismatches 427; Indels 175; Gaps 36;  
  
QY 3 LAALCRWGLLALLPPGAA---STQVCTGTDKMLRLPASPEHNLDMRLHLYGCGCOVVGCGN 59  
DB 7 LOVLC-----FLSLARGSEMGNQAVCPETTLNGLSVTGADNOYQTLVLYKCEVWQGN 62  
QY 60 IELTYLPTNVAISPLQDIOEVGVYLIANOVROVPLQRLIVRGTOLEFDNYALAVLDN 119  
DB 63 LEIVLTGHNADLSFLQWIREVTGYLVANRERSVLPPLRLAVRGVQVYDGKFAIPVM-- 120  
QY 120 GDPPLNNTPTVTCASRGGLRELOLSLTLLKGGVLIQRNPOLCYQDTILMKQIFPKNNQL 179  
DB 121 ---LNYNT---NSSHALRQLKFTQLTBLISGGVYIEKDKICHMDTIDMRIVRV-- 170  
QY 180 ALLTIDTRSRACHCSPKCGSRGCMGSSSSQSLTRIVCAGGC-ARCKGLPTDCCH 238  
DB 171 GAEIVKRNKNGANCPCPCHEVCCKG-RKMGREPDQCQLITIKICAPQCKGRCFGNPNQCHD 229  
QY 239 QCAAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTFESMNPBGRYTFGASCYTA 298  
DB 230 ECAGGCGSPDPTDCPCACRRFNDGACVPRCPFLVYNTLFTQLEBNPHTKYGGCVVAS 289  
QY 299 CPYNTLSTDVGSCTIVCPLEHNOEYTAEDGTQRCCKSCSPCARVCTGL--GMQYIANSKF 356  
DB 290 CPHNFV-VDQTFVCVARCPDRIKQEVYD-KHGLKMKCEPCGGLCPACACGTGSGSYQVYDSSN 347  
QY 357 IGITELFAGCKKIFGSLAFLPESFDGDPASNTATLQPEQQLVFTLEITGYLVISAMP 416

Db 348 ID----GFUNCTKILGNDLFDITLGLANDPMHKIPALDEKLNVRTRVEITGYLNIQSWP 403  
 Qy 417 DSLPDLVSFONLQVIRGRILNNGAVS-LTLOGLGISWLGSLRLSGLALIHNNHILC 475  
 Db 404 PHMNFVSFSLTITIGRSLVNRGFSLLIMKNLWVTSIGFSSLSKESISGRVYISANOQLC 463  
 Qy 476 FVHTVPMDOLEFRFNNFVSWLRVPKVSASHLE-----ECVGEGLACHOLCARGHCNCP 529  
 Db 464 YHHSILNTRL-----LRGFSBERLDIKYDRPLGECLAGKVCPLCSGGCGMP 512  
 Qy 530 GPTQCVNCSOFLRGOECVEBCRVLOGLPREYVNAHCLCPHEQOPONGSTTCGPRENDQ 589  
 Db 513 GPQGLCSKRNYSRGCVVTHCNFLOGEPRFVHAQCSCHPECLMEGTSTCGSGSDA 572  
 Qy 590 CVACAHYKDPFCVAVCPSPGVKPDLSYMPKFPDEEGACOPCEINCHSC--VDLDDKG 647  
 Db 573 CARGAHFRDGPCHCNSCPHILG--AKGPIYKYPDQNECHPCHENCOCNGEPLDCL 630  
 Qy 648 CPAEQRASPLTSVSAVVGILLVVGVVFGILIKRROQKR-KYMRRLLOETELVEPL 706  
 Db 631 GOAEVLMKSHPLVIAVVG--LAVILMILGSSFLYMGRRIRQNKRAMRYLERGESLEPL 688  
 Qy 707 TPGCAMPNOAMRLKETELRKVAVLGSAGFVYKGIWIPDGENVKIPVAIKYLRENTS 766  
 Db 689 DPS-EKANKVLAIRFKETELRKVLGSVGVVHKGIWIPESGSIKIPVICIKYIDKSG 747  
 Qy 767 PKANKELIDEAYVAVGSPVSRLLGICLTSTVQLVTLMPYCCLDHVENRGLRQSG 826  
 Db 748 RSFGQAVTDMHVLAVGSLDHAIYVRLGLCPSSIQVLTQVYPLGSLDHVQHEITLQ 807  
 Qy 827 DLNMCQIAKMSYLEEDVRLVHRDLARVNVKSPNHVKITDFGLARLLDIDETEHAD 886  
 Db 808 LLNMGVQIAKGVYLEHSHVHRDLALRVNMLKSPQOVADGVADLLPPDDKQLHS 867  
 Qy 887 GKVPIKMMALLESILRRFTHSDVMSGYVWVMEKMPGAKPYGICARAREIPDLLEGER 946  
 Db 868 EAKTPIKMMALLESILRRFTHSDVMSGYVWVMEKMPGAKPYGICARAREIPDLLEGER 927  
 Qy 947 LPQPICTIDVYIMVCMWIDSECRPRFELVSEFSRMAADPQRFVYIQNEDLGPASPL 1006  
 Db 928 LAQPOCTIDVYIMVCMWIDSECRPRFELVSEFSRMAADPQRFVYIKRAS-GPQTP- 985  
 Qy 1007 DSFFYSLEDDMDGLVDABEYVLPQOQFCPPAPGAGMHHRRSSSTRSGGDLT 1066  
 Db 986 -PAEFSVLTTEL-----QEALEPEL-----DLD 1010  
 Qy 1067 LGLEPSE-----EAPRSPLASEG-----AGSDVFDGDLGM 1098  
 Db 1011 LDLEAEELATLSGLASLPTGLTLTRPSQSLSPSSGMPNNGSLGEACLDASVLG 1070  
 Qy 1099 GAAKGLQSLPTHPSPLOQRYSEDPVLPSETDQV---APL-----TC-----SPQ 1142  
 Db 1071 GREQFSRPISLH-PIPRGR-----PASSESEHVTGSAELOEKVSVCRSRSSRSPR 1122  
 Qy 1143 PE-----YVNDPVRPQPPSPREGP-----LPABPAGATLERANTLSP-GRNGV 1186  
 Db 1123 PRGDSAYHSQRHSLTPVTPLSPGLEEDNGVYMDTLHRCASSREGTLSSVGLSSV 1182  
 Qy 1187 V-----KDVFAFGAVENPEVLTPOGGAPOHPHP 1216  
 Db 1183 LGTEEDED-----EVEYEMRKRRGSP-PRPP 1209

RESULT 11  
 EGFR\_DROME  
 ID\_EGFR\_DROME STANDARD; PRT; 1426 AA.  
 AC P04412; O61601; Q9W2G0; P81668;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)  
 (Gurken receptor) (Torpido protein) (Drosophila relative of ERBB).

GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 NX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).  
 RX MEDLINE=94350209; PubMed=8070664;  
 RA Clifford R., Schubach T.;  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 RT that several genetically defined classes of alleles cluster in  
 RT subdomains of the receptor protein."  
 RL Genetics 137:531-550(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Clifford R., Schubach T.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=85124611; PubMed=2982499;  
 RA Llywen E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;  
 RT "The Drosophila EGF receptor gene homolog: conservation of both  
 RT hormone binding and kinase domains."  
 RL Cell 40:599-607(1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Oregon-R; Tissue=Embryo;  
 RX MEDLINE=87002474; PubMed=3093080;  
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts."  
 RL Cell 46:1091-1101(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=99102120; PubMed=9882502;  
 RA Leeshkin A.M., Yu S.-Y., Katz J., Baker N.E.;  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila."  
 RL Dev. Biol. 205:129-144(1999).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RP STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga C., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,  
 RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burlingame K., Buzam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Rubeis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Durbin K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Schejter F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Welshbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [7]  
 RN SEQUENCE OF 959-1078 FROM N.A.  
 RP STRAIN=Daekwanryeong;  
 RC MEDLINE=85137938; PubMed=2983232;  
 RA Madsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;  
 RT "A *Drosophila* genomic sequence with homology to human epidermal  
 RT growth factor receptor.";  
 RL Nature 314:178-180(1985).  
 RN [8]  
 RN SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS.  
 RA MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shio B.Z.;  
 RT "Interallelic complementation among DER/Flb alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RN REVIEW  
 RP MEDLINE=97248481; PubMed=9094709;  
 RX Perrimon N., Perkins L.A.;  
 RA "There must be 50 ways to rule the signal: the case of the *Drosophila*  
 RT EGF receptor.";  
 RL Cell 89:13-16(1997).  
 CC -I- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MARK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMIOSEOSA  
 CC AND VENTRAL NEUROECODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -I- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -I- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANTAGEN OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AF052754; AAC08536.1; -  
 CC EMBL; AF052753; AAC08536.1; JOINED.  
 CC EMBL; AF052754; AAC08535.1; -  
 CC EMBL; AF052752; AAC08535.1; JOINED.  
 CC EMBL; K03054; AAAS1462.1; -  
 CC EMBL; K03417; AAAS1460.1; -

DR EMBL; K03416; AAAS0965.1; -  
 DR EMBL; K03418; AAAS1461.1; -  
 DR EMBL; AF109077; AAD26134.1; -  
 DR EMBL; AF109078; AAD26132.1; -  
 DR EMBL; AF109078; AAD26132.1; JOINED.  
 DR EMBL; AF109078; AAD26133.1; -  
 DR EMBL; AF109084; AAD26133.1; JOINED.  
 DR EMBL; AF109079; AAD26130.1; -  
 DR EMBL; AF109081; AAD26130.1; JOINED.  
 DR EMBL; AF109079; AAD26131.1; -  
 DR EMBL; AF109083; AAD26131.1; JOINED.  
 DR EMBL; AF109080; AAD26135.1; -  
 DR EMBL; AE003454; AAF46732.1; -  
 DR EMBL; X02293; CA26157.1; -  
 DR EMBL; X78930; CAAS5523.1; -  
 DR EMBL; X78918; CAAS5521.1; -  
 DR EMBL; X78919; CAAS5522.1; -  
 DR PIR; A00640; GQFE.  
 DR HSP; P11362; 1PKK.  
 DR FlyBase; FBgn003731; Egfr.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PD001009; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 7.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR TrEMBL; P55001; PROTEIN\_KINASE\_DOM; 1.  
 KW Tyrosine-protein kinase; Receptor; Phosphorylation; Transferase;  
 KW Developmental protein.  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 1 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 869 889 POTENTIAL.  
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 938 1198 PROTEIN KINASE.  
 FT NP\_BIND 944 952 ATP (BY SIMILARITY).  
 FT BINDING 971 971 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1063 1063 BY SIMILARITY.  
 FT MOD\_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 Query Match 28.1%; Score 1920; DB 1; Length 1426;  
 Blast Local Similarity 32.5%; Pred. No. 2.7e-95;  
 Matches 466; Conservative 184; Mismatches 430; Indels 356; Gaps 44;  
 QY 24 QVCTGTDMKRLPASPETHLDMLRLXGCGVQGNLELYPT-NASISPLDIOEYVG 82  
 DB KICIGTSKRLSVPSKHEHYNNLRDRYNTCTVQGNLTLTLPNPNLNDLFDNIREVY 159  
 QY 83 YVLAHNVQVPLQRLAIVGTOLF-----EDNYALAVLNDGDLNNTTPTVTSRGL 137  
 DB YILISHVDVKKVPEPKQIIRGRTLFSLVBEKXALFV-----TYSKM 203  
 QY 138 RELQRLSLTEILKGVLLQRPOLCYOOTLTKWDFHNGQNLALTLIDNRSRACHPSP 197  
 DB YTLLEPLDRLVNLQGVGHANNYNLCHMTTQMSSEIVSGTDAYNYVDTFAPRECPKHE 263  
 QY 198 MCKSRWGESSEDCQSLTRTVACGCA--RCKGPLPTDCHBQCACTGPKHSDCLAC 255  
 DB SCTHG-CWGEGRPKNCKFSKLTCSPOCAGRCYGPKECHLFCAGGCTGPTQDCLAC 322  
 QY 256 LHPHSGICELHCPALVTYNTDTESMNPGRRTYFGASCYTTACRYNTLSDVGSCTLYC 315  
 DB KNPFDEAVSKCECPMRKYNTPTTYVLETPBGKAYAGATVKECP-GHLRLDNGACVNSC 381  
 QY 316 PLHQEVYTAEDGTORCEKSKPCARVCGYGLGMQYIKANSKFIGITEL-----EPAGCKK 369

Db	382	PODKKDKGE	----	CVPCNGECPKTC	-----	PGVTVLHAGNIDSFNCTV	422		
Oy	370	IFGSLAFPLESEFDG	----	DPASNTA	-----	PLQPEOLQVFEETLEEITLYLISAMPSLDP	421		
Db	423	IDGNRIIDQTFSSGQDVYANTWMPRI	PLDPEREFVSTYKELITGLTINIEGTHPOERN	482					
Oy	422	LSVFONLOVIRGRILLHNGAY	-SLTLOGIGISWGLSRLRELGSLALIHNNHLCFVATV	480					
Db	483	LSYFNLLETIHGRQLMESMFALALIVKSSLSLEMRNLEKQISSGGSVVIQHNRLDCYSNI	542						
Oy	481	PMDOLEFRNNFTVSWLAVPKVSAHLEBCEVQEBGLAQHQLCARHGCMRGPRTQYCNCOF	540						
Db	543	RMPAIQKEPEQKV	-WVN-----	ENLRADLCXKNGTIGSDQCNEDCWMAGTDQCLTKCNF	596				
Oy	541	LRGQSECECRVLOQLPREYVWARHCLPCHPECCQONQSVTCFGEADQCVACAHYKDP	600						
Db	597	NFNGCTIADCGVISAAYK	-PDNRICKICHPECR-----	TONGAGADHCEQCAVHRDQ	648				
Oy	601	FCVARCP	-----	SGVK	-----	611			
Db	649	HCVSECPKMYKXNDRGVCRECHATCDGCTGPKDTIGICTGCTCNLAII	INNDATVKKCLKD	708					
Oy	612	----	PLSTMPITMKF	----	PDEBAGCP	-----	CPI-----	NCTH--	638
Db	709	DKCPD-GY	-FWEYVHPQEGSLKPLAGRAVCRKCHPLCELCTNYGHEOVCSKCTHYKR	765					
Oy	639	-----	-----	SC-----	VDLDDK	-----	647		
Db	766	REQCTECPADHYTDEBQRECFQRHPECNCGCTGPEADCKSCRNFKLFDANETGYPVNST	825						
Oy	648	-----	CPAEOR	-----	ASPLTS	-----	IYSAVAGILLVVL	673	
Db	826	MFNCTSKPBLEMRHNYQYTAIGPYCASPSPRSSKITANLDVNMFIITGAVLVPTICIL	885						
Oy	674	GVPFILLIKRQOKIRKYT	-MRLLQETELVEPLTPSGCAMPNOQMIRLKETELRKVY	731					
Db	886	CVV	-TYICRQKQKXKKEIVKTMALSCGESEBEPRLPANIGANLCKIRIVDAELRKGGV	943					
Oy	732	LGSGAFGVYGINWPDGENYKIPAIIVLRENTSPKANKELIDEAYMACGSPYVRL	791						
Db	944	LGMAGFVGNYGVMPEBENYKIPAIKELLKSTAESSEELREAYTMASEBENYNLKL	10030						
Oy	792	LGICLSTVOLVTOIMPYGCLLDHVENRGRIGSODLNMCMQIAGKSYLEDVRLVARD	851						
Db	1004	LAVCMSSQMLITQMLPGLCCLLDVYRNRDKIGSALLNMSTQIAGKSYIEEKRLVARD	10633						
Oy	852	LAARNVLKSPNHVKITDPGLARLIDIDETEHADGKVPRIKMALIESILRRFTHOSDV	911						
Db	1064	LAARVAVLQUTPSLVITDPGLAKLSSNSNEKKAAGKMPIKWALECTIRRVFTSKSDV	11233						
Oy	912	WSYGVTEMLTTFGAKPYDGI	PAREIPLLEKGEHLPPPTCTIDVYMIWAKCMMDISEC	971					
Db	1124	WAFGTVTWELLTFQGRPHENIPAKOIPOLIEVGLTEPELICSIDYCTLLSCMHLDAAM	11833						
Oy	972	RPRFRELVSFSRMARDPQREVIJONEDLG	--PASPLDSTYRSLSLEDD--	DMGDLVDA	10266				
Db	1184	RPTKQLTTFVAEFAFARDGKVIATIPGDKFTLPA	-----	YTSODEKDLIRKLAPTTDG	12366				
Oy	1027	EYVLVPOQGFCCPDPAFGAGMHHHRHSSSTRSGGDLTGLBSEBEAD	-----	RSP	10800				
Db	1237	SEALAKPDVYLOQKAPGPS	-----	HRTDCT	-----	DEMFKLNRKYCDP	1275		
Oy	1081	LAPBEGASDVDFG	----	DLGMAKAGLOS	LTTHDPSLQRISEDPYVLBETGYAPL	11377			
Db	1276	SNKQSSSTDDBDRDSAREVGNLR	-----	IDL	PVDEDDYILMP	13133			
Oy	1138	TCSPOPEYVNOPDYR	POPPSPREGPLPAARPAQATLERAKTLLSPQKGVVDVPAFGAV	11977					
Db	1314	TCQGGPNNNNMMN	-----	NPNQNNMAVGAAGM	-----	DLGVPYSV	1352		
Oy	1198	ENPEYVL	----	TPQGAAPQPH	-----	PPAPSP	AFDNL	YYWD	1230

```
Db      1353 DNEPYLANAQT/LGVGESPIPTQITGIPMGSGPMEXKVPMPGPSEPTSSDHEYND 1408
RESULT 12
ERBB_ALV ID_ERBB_ALV STANDARD: PRT: 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
OS Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OC Avian leukosis virus.
OX Virusess; Retroid viruses; Retroviridae; Alpharetrovirus.
RN NCBI_TaxId=11864; [1]
RP SEQUENCE FROM N.A. MEDLINE=85228222; PubMed=298784;
RX Nilssen T.W., Maroney P.A., Goodwin R.G., Rottman F.M., Crittenden L.B., Raines M.A., Kung H.-J.;
RA "c-erb activation in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino-truncated EGF receptor."; Cell 41:719-726(1985).
RL -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
-----
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TYFVLV.
DR HSPSP; P11362; IFKG.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SMO0219; TYKFC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene; Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BINDING 138 146 ATP (BY SIMILARITY).
FT FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MM; E705E33A0BE01FCC CRC64;

Query Match 25.6%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1,4e+86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
```



```

Db      59  NGSKTPSTIAAGVGGLLCLVAVGIGIYLRRR-HIVKRTLRLLQERLEVEPLTPSGE 117
Qy      712  MPNOQMRLKETELRKVKVLSGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANK 771
Db      118  APNOAHRLKETEFKVKVLSGSAFGTVYKGLWIPBEGKVKI PVAIKELREATSPKANK 177
Qy      772  EILDEAVYMAVGSPYVSRLLGICLTSTVOLVTOAMPYGCCLLDHRENRGRGSDLLNM 831
Db      178  EILDEAVYMAVDNPHVCRLLGICLTSTVOLVTOAMPYGCCLLDYIREHNDNIGSOYLNM 237
Qy      832  CMOIAKGSYLEDVYLVRDLAARVLYKSPHNVKITPFGRLRLDIDETEVHADGKVP 891
Db      238  CVQIAKGNVYLERLVLRDLAARVLYKTPQHVKITPFGRLAKLGADEKEVHADGKVP 297
Qy      892  IKMALESILRRRFTHQSDWVSYGTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQP 951
Db      298  IKMALESILRRITHQSDWVSYGTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy      952  ICTIDVYMIWKCMWIDSECRPRELVSFSPRMARDQRFVVIQ-NEDLGASPLDSTF 1010
Db      358  ICTIDVYMIWKCMWIDASRPKRELIAFSSKMARDPRLVIVIGDERHMLPSPDTSKF 417
Qy      1011  YRSLEDDDMGDLVDAAEYLVPOQGFPCPDPAAGAGVNHHRHSSSTRSGGDLTLGLE 1070
Db      418  YRTIMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1071  PSEBAPRSP-----APSEGASGVFPGDLGAKGLQSLPTHPSPLOKYSDDPVP 1125
Db      450  -----SRTPLLSLSATSNNSATNCID-----RNGQGHPRVEDSFVQYSSDPICN 495
Qy      1126  LPSET--DGVAAPLTCSPQEPQEVVNOPDVROPSPREBGLPARAGATLEAKTLSGCK 1183
Db      496  FLEESIDDFL-----PAPEVNO--LMPKKS-----TANVQ 526
Qy      1184  NGVVKDVE-----AFGAVENPEYLTPOGGAAPQHPAPSPAPDNLVY 1228
Db      527  NQIVNNISLTAISKLPMSRYONSHSTAVDNEYL-----NTNQSPLAKTVESSPY 578
Qy      1229  WDO-----DPEP-----RCAPSPSTFKCTPRAENPEYGLDVP 1260
Db      579  WIGSQNHQINLDNPDYQDFLENETKPNGLKVPAAENPEYLRVAP 625

RESULT 13
ERBB_AVIER STANDARD: PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
   gene family.";
RL Cell 35:71-78(1983).
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debutte B., Henry C., Benaisa M., Bisette G., Claverie J.-M.,
   Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
   new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC      tyrosine phosphate.
CC      -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC      ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC      IN CHICKENS.
CC      -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC      RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, K02006; AAA42394.1; ALT_INIT.
DR      EMBL, K01216; AAA42400.1; -.
DR      PIR: A00644; TVYU.
DR      HSSP, P11362; 1FGK.
DR      InterPro; IPR000719; Euk_pkinase.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00069; pkinase; 1.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00219; TyrKc_1.
DR      PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR      PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132..399
FT      NP_BIND 138..146 ATP (BY SIMILARITY).
FT      BINDING 165..165 ATP (BY SIMILARITY).
FT      ACT_SITE 257..257 BY SIMILARITY.
FT      CONFLICT 29..29 S -> W (IN REF. 2).
FT      CONFLICT 140..140 S -> F (IN REF. 2).
FT      CONFLICT 146..146 I -> V (IN REF. 2).
SQ      SEQUENCE 604 AA; 67633 MW; 76EBD067450609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4,1e-84;
Matches 360; Conservative 76; Mismatches 126; Indels 126; Gaps 16;

Qy      593  CAHYKDPFCVCARCSGVKPDLSYPIKFPDEBACQPCPINCTHSCVDLDDKGPAAO 652
Db      3  CAHFIDGPHCVAKCAGVYGENDTL-VRKYDANAVCOLCHPNCCTRGCKGPGLEGCP-- 58
Qy      653  RASPLTSIVAVV-GILVVVGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGA 711
Db      59  NGSKTPSTIAAGVGGLLCLVAVGIGIYLRRR-HIVKRTLRLLQERLEVEPLTPSGE 117
Qy      712  MPNOQMRLKETELRKVKVLSGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANK 771
Db      118  APNOAHRLKETEFKVKVLSGSAFGTVYKGLWIPBEGKVKI PVAIKELREATSPKANK 177
Qy      772  EILDEAVYMAVGSPYVSRLLGICLTSTVOLVTOAMPYGCCLLDHRENRGRGSDLLNM 831
Db      178  EILDEAVYMAVDNPHVCRLLGICLTSTVOLVTOAMPYGCCLLDYIREHNDNIGSOYLNM 237
Qy      832  CMOIAKGSYLEDVYLVRDLAARVLYKSPHNVKITPFGRLRLDIDETEVHADGKVP 891
Db      238  CVQIAKGNVYLERLVLRDLAARVLYKTPQHVKITPFGRLAKLGADEKEVHADGKVP 297
Qy      892  IKMALESILRRRFTHQSDWVSYGTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQP 951
Db      298  IKMALESILRRITHQSDWVSYGTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy      952  ICTIDVYMIWKCMWIDSECRPRELVSFSPRMARDQRFVVIQ-NEDLGASPLDSTF 1010
Db      358  ICTIDVYMIWKCMWIDASRPKRELIAFSSKMARDPRLVIVIGDERHMLPSPDTSKF 417
Qy      1011  YRSLEDDDMGDLVDAAEYLVPOQGFPCPDPAAGAGVNHHRHSSSTRSGGDLTLGLE 1070

```



```

Db 418 YRTLMEEEDMDIADADYLVPHOGFF-----NSPST----- 449
Qy 1071 PSEEEARSPV-----APSEGASDVDFDGLGMAKGLQSLPTHDPSPLORYSEDPTVP 1125
Db 450 -----SRTPLLSLSATSNNNSATNCID-----RAGQGHPRVREDSFVGRYSDDPCPN 495
Qy 1126 LPST--DGYVAPLTCSQPEYVNOQDVPRPQPSRECPPLPAPRAGATLERAKTLSFGK 1183
Db 496 FLESIDGFL-----PAPYVNO--LMPKKPSTAM----- 524
Qy 1184 NGVVKDVFVAF-----GGAVENPEYLTTPCGGAAPQHPPEAPSPAFD 1224
Db 525 --VNOQIYNFISLTAISKLPMDSKYQNSHSTAVDNEPYL-----NTNOSPILAKTYFE 574
Qy 1225 NLVYWDQDPPERGAPPTFKCTPTAENPEY 1254
Db 575 SSPYMIOSGNHQ-----INLDNPDY 594

RESULT 14
ERBB_AVIEU STANDARD; PRT; 540 AA.
ID_ERBB_AVIEU
AC P11273;
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (BC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retrocid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RA "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759 (1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL, M13179; AAA42401.1; -.
DR PIR, A25231; TVEVEB.
DR HSSP, P11362; IEGK.
DR InterPro, IPR000719; Euk_Pkinase.
DR InterPro, IPR001245; Tyr_Pkinase.
DR Pfam, PF00069; Pkinase; 1.
DR ProDom, PD000001; Euk_Pkinase; 1.
DR SMART, SM00219; TYRKC; 1.
DR PROSITE, PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE, PS00109; PROTEIN_KINASE_TYR_1.
DR PROSITE, PS50011; PROTEIN_KINASE_DOM_1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN_KINASE.
FT ND BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 23.8%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 2.9e-80;

```

```

Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
Qy 593 CAHYNDPFCVACRCGKVPDLSTVPIWKFPDEBGAOCPCINTGHSQVLDLQKGPAPQ 652
Db 3 CAHFDGHCYKACGACGAGVLGENDTL-VKRYADANAACVQLCHNCNCRGCGPGLGECF--- 58
Qy 653 RASPLTISVAVV-GILLVVLGVVFGILIKRROOKIRKRYTMRLLQETELVEPLTPSGA 711
Db 59 NGSKTPSIAGAVGGLLCLVYVGLIGLYLRR-HIVKRTLRLLQERLELVEPLTPSGE 117
Qy 712 MPNOQMRLKETELARKKVLGSGAFGVYKCIWIPDEBNKIPLAIKYLRENTSPKANK 771
Db 118 APNOAHLKLEKETEKVKVGFAGFTVYKGLMPEBEKVTIPLAIKELRENTSPKANK 177
Qy 772 EILDEAYMAGVSGSVYSLRIGICLTSTVOVLTOLMPYGCCLDHRNRCGLSDGLNW 831
Db 178 EILDEAYMASVDNPHVCRLLGICLTSTVOLTLQMPYGCCLDVRHNDNIGSQYLLNW 237
Qy 832 CMOIAKGNSTLEDVRLVHRDLAARNLVKSPHNVKITDFGLARLLDIDETEHADGKVP 891
Db 238 CVQIAKGNVLEBRMVRDLAARNVLKTPQHVKITDFGLAKOIGADEKEVHAEGKVP 297
Qy 892 IKMALESILRRRTHQSDVWSYGVTTWELMTFGAKPYDGIPIAREIPDLLEKGERLPOP 951
Db 298 IKMALESILHRIYTHQSDVWSYGVTTWELMTFGSKPYDGIPIASEISVLEKGERLPOP 357
Qy 952 ICTIDVYIMVKNWIDSECPRELYSEFSRMAPDQRFVVO-NEDLGASPLDSTF 1010
Db 358 ICTIDVYIMVKNWIDSDSRPKRELTAESKKAARDPRRLVIGDERMHLPSPTDSKF 417
Qy 1011 YRSLEDDMDGLVDAEYLVPOGFFCPDPAPGAGVHHRHRSSTFRSGGDLTLGLE 1070
Db 418 YRTLMEEEDMDIADADYLVPHOGFF-----NSPST----- 449
Qy 1071 PSEEEARSPV-----APSEGASDVDFDGLGMAKGLQSLPTHDPSPLORYSEDPTVP 1125
Db 450 -----SRTPLLSLSATSNNNSATNCIDRNG-----H----- 476
Qy 1126 LPSTNDGYVAPLTCSQPEYVNOQDVPRPQPSRECPPLPAPRAGAT-LEAKTISFGK 1184
Db 477 -PVREDFL-----PAPYVNO--LMPKKPSTAMVNOQIYNFISLTAISKLPIDSRVON 527
Qy 1185 GVVDQVFAFGAVENPEYL 1203
Db 528 -----SHSTAVDNEPYL 539

RESULT 15
EGFR_CHICK STANDARD; PRT; 703 AA.
ID_EGFR_CHICK
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (CER
DE (Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RA "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND

```

CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
CC dimerization, internalization of the EGF-receptor complex,  
CC induction of the tyrosine kinase activity, stimulation of cell DNA  
CC synthesis, and cell proliferation.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M20386; AAA48760.1; .  
DR Interpro: IPR000494; EGFR\_L\_domain.  
DR Interpro: IPR000719; Euk\_Pkinase.  
DR Interpro: IPR002174; Purin-like.  
DR Interpro: IPR001245; Tyr\_Pkinase.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR SMART: SM00261; FU; 4.  
DR PROSITE: PS00107; PROTEIN KINASE ATP; PARTIAL.  
DR PROSITE: PS00109; PROTEIN KINASE\_TYR; PARTIAL.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; PARTIAL.  
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT CHAIN 1 30  
FT DOMAIN 31 654  
FT TRANSMEM 655 667  
FT DOMAIN 668 667  
FT DISULFID 197 206  
FT DISULFID 201 214  
FT DISULFID 222 230  
FT DISULFID 226 238  
FT DISULFID 239 247  
FT DISULFID 243 255  
FT DISULFID 258 267  
FT DISULFID 271 298  
FT DISULFID 302 314  
FT DISULFID 318 333  
FT DISULFID 336 340  
FT DISULFID 513 522  
FT DISULFID 517 530  
FT DISULFID 533 542  
FT DISULFID 546 562  
FT DISULFID 565 581  
FT DISULFID 589 589  
FT DISULFID 592 601  
FT DISULFID 605 627  
FT DISULFID 630 638  
FT DISULFID 634 646  
FT CARBOHYD 134 134  
FT CARBOHYD 190 190  
FT CARBOHYD 200 200  
FT CARBOHYD 359 359  
FT CARBOHYD 368 368  
FT CARBOHYD 420 420  
FT CARBOHYD 420 420  
FT CARBOHYD 573 573  
FT CARBOHYD 578 578  
FT CARBOHYD 613 613  
FT CARBOHYD 633 633  
FT CARBOHYD 648 648  
FT NON\_TER 703  
SQ SEQUENCE 703 AA; 77427 MM; AFE2DE11B735A690 CRC64;

Query Match 23.0%; Score 1571; DB 1; Length 703;  
Best Local Similarity 43.8%; Pred. No. 5.Se-77;

Matches 313: Conservative 113: Mismatches 252: Indels 36: Gaps 15:  
QY 8 RWGLLALLPPGAA-----STVCTGTDMLRLPASEPTHLDMLRHLYOCGVQGNLE 61  
DB 13 RGAAVLVLLGLVALCSAVERKVCQGTNNKLTOLGHNVEDHTSLQRYMNCNEVLSNLE 72  
QY 62 LTYLPTNASLFLDIOGVGVLIJAHNOVPLQRLRYRGTOLFEDNVALAVLNDG 121  
DB 73 ITVEHNRDLFTFLKIQEAVGVLIJALNMVDVIFLENLQITRGVNLVINDSPALAVSNH 132  
QY 122 PLNNTPVTGASPGGLRELQURSLTEILKGVLIQRPOLCYODTILMKDIFHKNOQLAL 181  
DB 133 -MNTQ-----GLRELPMKRSEILNGGVKISNNPKLQCMQDVLNMNDIDTSRK-PL 182  
QY 182 TLID-TNRSRACHPCSPMKCKSRCKGSESSDQSLTRVYAGGCA-RCKGPLPTCCHEQ 239  
DB 183 TVLDFASVLSGCPKCHPCTEDHCHGAGEQNCQTLTKVICAQCCSGRGKVPKPSDCHNQ 242  
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHCPALVNTDTPFESMPREGRTFGASCVTAC 299  
DB 243 CAAGCTGRESDBCLACKRFDDATCTKCPPLVLTPTTYQMDVNBKYSFGATCVREC 302  
QY 300 PYNVLTSDVGSCTLVCPILHNOEVTAEQTCCKSKPCARVCYGLGMQYIKANSKFIGI 359  
DB 303 PHNVVTDHGS CVRSCNTDTEV-ENGVKCKCKDGLCSKVCNGIGIGELKGIUS-INA 360  
QY 360 TELE-FACCKKIFPSGLAFPSFGDPASNTAPLOPELOVPELTLEITGYLYISAWPS 418  
DB 361 TNISFKQCTKINDVSLPVAFLGDAFTKTLPLPKLDVFRVKEISGFLLIQAMPDN 420  
QY 419 LPDLSFQNLQVIRGRIHNGAYSLTLOGLGISWGLSRLEISGLALJHNNHLCVPH 478  
DB 421 ATDLVAFENLEITIRGTRKQGOYSILAVNLKIOSLGRLSEISDGIJALMKKNLCYAD 480  
QY 479 TVPMDOLFRFNNFTVSEFWLRVPKVSASHSECEVGEGLACHQLCARGHCMGPGPTCVACS 538  
DB 481 TMNRSLEFATQS-----QTKKIIONRNKNDCTADRHVCDPLCSGVGCMGPGFHCFSR 534  
QY 539 QFLGQSECEVCECVLOGIPREYVNAHRLCPHPECQPNG--SVTCGPEADQCVACAH 595  
DB 535 PFSRQKECVKQCNILQSEPREFERSKCLPCHSECLVQNSTAYVNTCSGPGDHCMCAH 594  
QY 596 YKDPFPCVARGPSGVKPDLSYMPIMKPPDEGACQPCINCHSCVDDDDGGCPAEGQAS 655  
DB 595 FIDPHCVKACPAVLGBNDTL-VKXTADANAVCOLCHPNCCTRGKGGLECCP--NGS 650  
QY 656 PLTSIVSAV-GLLVVLGVVFGILIKRQOKIRKTYMRLLOETELVEPLTP 708  
DB 651 KTSIAGVGVGGLCLVVGIGLYLARR-HIVRKRLRLRLQERLVEPLTP 703

Search completed: July 22, 2003, 08:46:40  
Job time : 20.2793 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-465-479-14

Perfect score: 6842  
Sequence: 1 MEALALCRWGLLALLPPGA.....TFKGTPTAENPEVIGLDPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_prodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6125	89.5	1259	6	O18735
2	3113	45.5	1209	11	O9QX70
3	3084	45.1	1210	11	O9EP98
4	2719	39.7	1165	13	O9YH40
5	2662.5	38.9	1137	13	O9W6F6
6	2280	33.3	1328	13	P79754
7	2009.5	29.4	1433	5	O9BIF9
8	1871	27.3	419	4	O9UK79
9	1739	25.4	367	11	O8R2X1
10	1720	25.1	729	15	O86712
11	1718	25.1	567	15	O86714
12	1697.5	24.8	412	4	O8WYV0
13	1653.5	24.2	962	15	O64895
14	1645	24.0	545	15	O85468
15	1475.5	21.6	655	11	O9WVFS
16	1459.5	21.3	643	11	O9ERV6

17	1275	18.6	1193	5	O9Y1X8
18	1180.5	17.3	1368	5	O23821
19	1167	17.1	1717	5	O26566
20	1108	16.2	527	13	O9O836
21	1002	14.6	478	11	O9ESR0
22	924.5	13.5	599	13	O9PSH2
23	906	13.2	165	4	O14256
24	806.5	11.8	346	13	P11776
25	797	11.6	176	11	O923V5
26	778	11.4	435	5	O8SZM1
27	754.5	11.0	311	13	O9P162
28	734.5	10.7	1362	13	O9PV24
29	734	10.6	331	4	O9BUD7
30	723	10.7	149	6	O9BG66
31	721	10.5	1671	5	O9UJVS
32	694	10.1	1418	13	O93457
33	690.5	10.1	1368	13	O8UW85
34	669.5	9.8	1432	13	O8UW86
35	667	9.7	1412	13	O8UW84
36	666.5	9.7	1472	5	O9U5A8
37	654	9.6	1358	13	O73798
38	644.5	9.4	1418	13	O8UW83
39	632	9.2	1245	13	O9YGH8
40	631.5	9.2	2144	5	O9VD94
41	625	9.1	1371	11	O9QVW4
42	598	8.7	987	11	O91VW0
43	593	8.7	987	11	O99WR2
44	591	8.6	935	4	O9EL35
45	587.5	8.6	1036	4	O07912

## ALIGNMENTS

RESULT 1  
ID O18735 PRELIMINARY; PRT: 1259 AA.  
AC O18735;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Erdb-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
NCBI\_taxid=9615;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP Yokota H.;  
RA "CDNA cloning of erdb-2 from canine mammary gland";  
RT Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1; -  
DR HSSP; P1362; 1FGK.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR00494; EGFR L domain.  
DR InterPro: IPR00719; Euk\_Pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_Pkinase.  
DR InterPro: IPR004019; YLP motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; Pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_Pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR ATP-binding; Transferase; Tyrosine-protein kinase.  
KW SEQUENCE 1259 AA; 137989 MW; E37364D9C4ACD46 CRC64;

O9Y1X8 ephydatia f  
O23821 caenorhabdi  
O26566 schistosoma  
O9O836 gallus gall  
O9ESR0 rattus norv  
O9PSH2 gallus gall  
O14256 homo sapien  
P11776 xiphophorus  
O923V5 rattus norv  
O8SZM1 drosophila  
O9P162 xiphophorus  
O9PV24 xenopus lae  
O9BUD7 homo sapien  
O9BG66 oryctolagus  
O9UJVS biophalari  
O93457 scophthalmu  
O8UW85 paralicthy  
O8UW86 paralicthy  
O8UW84 paralicthy  
O9U5A8 bombyx mori  
O73798 xenopus lae  
O8UW83 paralicthy  
O9YGH8 scophthalmu  
O9VD94 drosophila  
O9QVW4 rattus sp.  
O91VW0 mus musculu  
O99WR2 mus musculu  
O9EL35 homo sapien  
O07912 homo sapien

Query Match 89.5%; Score 6125; DB 6; Length 1259;  
 Best Local Similarity 89.6%; Pred. No. 0;  
 Matches 1134; Conservative 44; Mismatches 76; Indels 12; Gaps 3;

QY 1 MELAALRWGILLALLPFGAASSTOYCTGTDKMLRLPASPETHLMDLRLYGGCOVQGNL 60  
 DB 1 MELAAWRWGILLALLPFGAAGTGVCTGTDKMLRLPASPETHLMDLRLYGGCOVQGNL 60  
 QY 61 ELFTLPNASTLSFQDIOEVGVYVLIANOVROYVLOBLRLVIRGTOLPEDYVALAVDNG 120  
 DB 61 ELFTLPNASTLSFQDIOEVGVYVLIANSQVROJPLQRLRLVIRGTOLPEDYVALAVDNG 120  
 QY 121 DPLNNTPTVGTASFGILRELIQSLTEILKGVLIQRNPOLCYQDTILMKDIFKNNOLA 180  
 DB 121 DPLEGGIPARCAAGGIRELIQSLTEILKGVLIQRNPOLCHQDTILMKDIFKNNOLA 180  
 QY 181 LTLIDTRSRACHSCSPMKCKSRWGESSEDCOSLTRVVCAGGACRCGLPTCCHEQC 240  
 DB 181 LTLIDTRSRACHSCSPMKCKSRWGESSEDCOSLTRVVCAGGACRCGLPTCCHEQC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300  
 QY 301 YNVLSTVGSCTVCPHNOEVTADGTORCEKCSKPCARVCYGLMQYIKANSKFIGIT 360  
 DB 301 YNVLSTVGSCTVCPHNOEVTADGTORCEKCSKPCARVCYGLMQYIKANSKFIGIT 360  
 QY 361 ELEFAGCKKIFGSLAFIPESFDGPASNTAPLOPEQLOVEFTLEITGYLYISAMPDLP 420  
 DB 361 IOEPAGCKKIFGSLAFIPESFDGPASNTAPLOPEQLOVEFTLEITGYLYISAMPDLP 420  
 QY 421 DLASFQMLQVIRGILHNAYSLTLQGLISMLGLRSLREBSGSLALHNTHLCFYNTV 480  
 DB 421 NLSVFOULRVIRGVLHDGASLTLQGLISMLGLRSLREBSGSLALHNTHLCFYNTV 480  
 QY 481 PMDOLFENFTVSFWLIRPKVASASHLEECVGEGLACHQLCARHCWGPFGTCVNSQF 540  
 DB 481 PMDOLFENFTVSFWLIRPKVASASHLEECVGEGLACHQLCARHCWGPFGTCVNSQF 540  
 QY 541 LRQECVEECRVLOGLPREVYNARHCLPCHPECOPONSVCPCFPREADQCAACHYDPR 600  
 DB 541 LRQECVEECRVLOGLPREVYNARHCLPCHPECOPONSVCPCFPREADQCAACHYDPR 600  
 QY 601 FCVARCPSGVKPDLSYMPIMKFPDEBAGACOPPCINCHSCVDLDDKCGPABQASPLTSI 660  
 DB 601 FCVARCPSGVKPDLSYMPIMKFPDEBAGACOPPCINCHSCVDLDDKCGPABQASPLTSI 660  
 QY 661 VSAVVGILLVVLGVVFGLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRI 720  
 DB 661 VSAVVGILLVVLGVVFGLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRI 720  
 QY 721 LKETELRKVKVLGAGFGTYVKGIMIPGENVVKI PVALIKVLRNTSPANKELLIDEAYVM 780  
 DB 721 LKETELRKVKVLGAGFGTYVKGIMIPGENVVKI PVALIKVLRNTSPANKELLIDEAYVM 780  
 QY 781 AGVGSFYVSRLLGLCLTSTVOLVTLMPYGCILDHVENRGLSGODLLNMCQIAGKMS 840  
 DB 781 AGVGSFYVSRLLGLCLTSTVOLVTLMPYGCILDHVENRGLSGODLLNMCQIAGKMS 840  
 QY 841 YLEEDVRLVHRDLAARNVLVKSPPNHVKITDFFGLARLDIDETFYADGKVPDKMALESI 900  
 DB 841 YLEEDVRLVHRDLAARNVLVKSPPNHVKITDFFGLARLDIDETFYADGKVPDKMALESI 900  
 QY 901 LRRRFTQSDVWSVGVTVMELMTGAKPYDGI PAREIPDLLEKEERLPORPICIYDVM 960  
 DB 901 LRRRFTQSDVWSVGVTVMELMTGAKPYDGI PAREIPDLLEKEERLPORPICIYDVM 960  
 QY 961 MVKCMWIDSECRPFREILVSEFSMAADPORFVVIIONEDLGPASPLDSTFYRSLLEDDM 1020  
 DB 961 MVKCMWIDSECRPFREILVSEFSMAADPORFVVIIONEDLGPASPLDSTFYRSLLEDDM 1020  
 QY 1021 GDVDAEYLVLPQGGFCFPDPAFGAGMWHHRSSSTRSGGDLTLGLSEBEARPSR 1080  
 DB 1021 GDVDAEYLVLPQGGFCFPDPAFGAGMWHHRSSSTRSGGDLTLGLSEBEARPSR 1080

DB 1014 GDVDAEYLVLPQGGFCFPDPAFGAGTAHRRSSSTRNGGELTLGLEPSEEPKSP 1073  
 QY 1081 LAPSEAGSDVFDODLDMGAAGKQSLPTNHPSLQRYSEPTVPLPSETGGYVAPLICS 1140  
 DB 1074 LAPSEAGSDVFDODLDMGAAGKQSLPSODPSLQRYSEPTVPLPSETGGYVAPLICS 1133  
 QY 1141 PPEYVNPQDVRPQPSPREGLPAPAPAGATLER-----AKTLPCKNGVVKOVFAFG 1195  
 DB 1134 PPEYVNPQDVRPQPSPREGLPAPAPAGATLER-----AKTLPCKNGVVKOVFAFG 1193  
 QY 1196 AVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGADPSTFKGTPAENPEYL 1255  
 DB 1194 AVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGADPSTFKGTPAENPEYL 1253  
 QY 1256 GLDVPV 1261  
 DB 1254 GLDVPV 1259

## RESULT 2

Q9GX70 PRELIMINARY; PRT; 1209 AA.  
 AC Q9GX70;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor.  
 GN EGR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RX MEDLINE=9025688; PubMed=234246;  
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
 RA Eard H.S.;  
 RT "A truncated, secreted form of the epidermal growth factor receptor is  
 RT encoded by an alternatively spliced transcript in normal rat tissue."  
 RL Mol. Cell. Biol. 10:2973-2982(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Petch L.A.;  
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Guttridge K., Dawson T.L., Eard H.S.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; M37394; AA014008.1; -.  
 DR HSSP; P1362; IFGK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_DKinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_DKinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_DKinase; 1.  
 DR SMART; SM00261; Fu\_3.  
 DR SMART; SM00219; TykC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE87F6CC187773 CRC64;

Query Match 45.5%; Score 3113; DB 11; Length 1209;  
 Best Local Similarity 49.9%; Pred. No. 6; Je-227;

Matches 641; Conservative 167; Mismatches 353; Indels 124; Gaps 30;

```

Qy 3 LAALCRWGLLALLPPGA--ASTOVCTGDMKRLRLPASBETHLDMRHHYOGQVQGMLE 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 LAALCAAG-----GALBEKKVCGQTSNRLTQGTDEHDLISQRRFNNEEVLGNLE 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 62 LTYLPTNASLSFLDIOEVQGVYLIAHQVROVPLQRLRYRGTOLEFEDNALAVLNGD 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 ITTYQARNYDLFLKTIQEVAGVYLIALNTVERIIPLENQIIRGNALYENTVALAVLSN-- 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 122 PLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQRNPQLCYQDITLWKDIPIKNNQAL 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 -----YGNKKGLRELPMKNLOEILIGAVRFNNPILCMETIIQMRDIV-QVVFISN 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 182 TLIDTNRS-RACHPCSPMKSGSRCESESDQSLTRTYCAGGCA-RCKGPLPTCCHEQ 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 MSMVQGRHLTGCPKCDPSCPGNSCWRGEBENQKLTIKIICAQGCRGRGRSPSCCINQ 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 240 CAAGCTGPRHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 CAAGCTGPRSDCLVCHRFDEATCKDTCPLMLYNTTYQMDVNPGEYSGATCVKCC 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 300 PNYLTSTVGSGCTLVCPHNOEVTAEQTORCEKSKCARCYGLGMOYIKANKFTIGI 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 PRNVVVDHSGCVRACGPDYEV-BEDVSKCKKCDGCFKRCVNGIIGEFK-DTLSTINA 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 360 TELE-FACCKKIFGSLAFLEPSFDGDPASNTAPLOPEOLOVEETLEETLYLYIACMPD 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 TNIHFKKYCTAISGLHLPLVAFKDSFTTRPDLPRELEIKYKEITGFLIIMABEN 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 419 LPDISVFONLOVIRGRILHNGASYSLTQGLGSIWGLSRLSRLSGSLALIHNTLCEVH 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 WTDIAFENLEIIRGRTRKHQGFSLAVVGLNITSLGLSLKEISDGVYISGNRLCYAN 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 479 TVPMDOLFRFNNFTYSFWLAVPKVSASHLE-ECVEGELACIQLCARGHCKMGRPGVCNC 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 TINMKKLGSTPN-----QTKIMNNAEKDCATNHCPLCSSECGEPEPTDVCSC 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 538 SQPLRGQCEVECEVLTQGLPREYVNAHCLPCHPEQONGSVTFCGEADQCVACAYK 597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 QNVRGRRCVCKNMLEBGEPRFVENSEICHPCELCPTQNMNITGTGPRNCICAHYV 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 598 DRPCCVARCEBGCVPDLSTYMPYKFPDEEGACQPCPCINCTHSCVDLDKXCEPAEGRASP- 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 DGPHCVKTCPSGIMGENNTL-VMKFADANNVCHLCHANCTYGCAGPGLKGC--QQPEGPX 643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 657 LTSIVSAVAGILVAVGVGVFI-LIKRQOKIRKYMRRLLOETELVEPLTPSGAMNQ 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 IPSIATGIVGGLFIV-VALGICLPMRRQDLVRKTRRLLOERELVEPLTPSGEAPNQ 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 716 AQMRILKETELRKVYLGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANKETILD 775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 703 AHLRLKETEFKIKVYLGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANKETILD 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 776 EAYMAGVGSPPYVRLGICLTSTVQVLTQMLPYGCLLDHVENRNGRLGSDLLMNCQI 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 EAYMASVQDNPHVCRLIGICLTSTVQVLTQMLPYGCLLDYVERHKNIGSQYLLMNCQI 822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 836 AKGWSYEDVYLVHRDLAARVVLKSPHVKITDPCGLARLIDIBETVHAAGKVPYKMM 895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 AKGNMYLEDRLVHRDLAARVVLKSPHVKITDPCGLARLIDIBETVHAAGKVPYKMM 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 896 ALBSILRRPTHQSDVMSYGTVMELMFGAKPYDGI PAREIPDLLEKGERLPORPICI 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 ALBSILRRPTHQSDVMSYGTVMELMFGSKPYDGI PASISISLLEKGERLPORPICI 942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 956 DVTYIMYKCMWIDSECPRELYSESRMARDPQRFVIO-NEDLGPASPLDSTFYKSL 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 DVTYIMYKCMWIDSRKRELLLEFSKMARDPQRFVIOGDERMHLPSPTDSNFYAL 1002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1015 LEDDDMGDLVDAEEYLVQOGFPCPDPAKAGMWHHRHSSSRSGGDLTLGLEPSEE 1074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 MEEDMEDVDADAEYLIPQSGF-----NSPST----- 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 1075 BAPRSLPABEGAGDVDFDGLGMGAAGLQSLPTHDSPLQRYSEDPVLPSPER--DG 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1031 --SRTPSLSLASNSN-----SSTVACINRNSCRVKEDAFLQRYSDSPFSLTEIDIDT 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1133 YVAPLTCSPPQPEYVQNPDPVRPOPSPREGPLPAAPRAGATLERAKTSLFGKNGVVKDVA 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 FL-----PVPEYINQ-SVPRKPRAGSVQNPVYHNPPLHP-----APGRDLHYQN--P 1127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1193 FGAAVENPEYL-TPCGGAAPQHPHPPAFSPADNLVYNDQ-----DP-----PE 1235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1128 HSNAVSNPEYLNMTAQ-----PTCLSSGFSLSALMIQKSHQMSLDNPDYQODFFPK 1178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1236 RGAPPSTFKGTPTAENPEYGLDVP 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1179 EAKPRGIFKG-PTAENAEYLAVAP 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 3

```

Q9EP98 PRELIMINARY; PRT: 1210 AA.
ID Q9EP98
AC Q9EP98:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC.
RA Reiter J.L., Threagill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sincalir C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mañlle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RN [2]
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threagill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schehl C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mañlle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAC28045.1; -.
DR EMBL; AF275364; AAC28045.1; JOINED.
DR EMBL; AF275365; AAC28045.1; JOINED.
DR EMBL; AF275367; AAC24386.1; -.
DR HSSP; P11362; IFGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; Cytc heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Puri-like.
DR InterPro; IPR002290; Ser_thr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PFO1030; Recep_L domain; 2.
DR PRINTS; P00109; TYRKINASE.
DR PRODOM; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Tyrc; 1.

```

DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN\_1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 KM ATP-binding; Receptor; Transferase.  
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 45.1%; Score 3084; DB 11; Length 1210;  
 Best Local Similarity 49.4%; Pred. No. 1e-224;  
 Matches 632; Conservative 169; Mismatches 359; Indels 120; Gaps 27;

QY 11 LLLALLPFGAA--STVCTGTDMKRLRASPETHLMDLRHLYOGCOVYQGNLELTLYPTN 68  
 DB 14 LLLALLPFGAA--STVCTGTDMKRLRASPETHLMDLRHLYOGCOVYQGNLELTLYPTN 73  
 QY 69 ASLSFLDIOEVQGVVLIHNOVQVPLQRLIYRGVTLQFEDNVALVLDGDLNATTP 128  
 DB 74 YDLSFLKTIQEVAGVLIHNOVQVPLQRLIYRGVTLQFEDNVALVLDGDLNATTP 124  
 QY 129 VTGASPGRLRELQRLSLTEILKGVLIQNPOLCYODTILMKDI----FKNNQLATLI 184  
 DB 125 YGNNRGLRLRLPRLNLOEILIGAVRFNNPLCLMDITIQMRDLYQVNFNMSMDL--- 180  
 QY 185 DTNRSRACHPCSPWCKSGRSGESSEDCQSLRTVYAGCA--RCKPLPTDCHEQCAAG 243  
 DB 181 -QSHPSGCPKCDPSCPNCSGCGGEGNCKLTKIICAQCGSHRCGRGSPDCCHQCAAG 239  
 QY 244 CTGPKHSQCLACHPHNSGICELHCPALVTYNTDTFESMPREGRYTTGASCATAPVNY 303  
 DB 240 CTGPKHSQCLACHPHNSGICELHCPALVTYNTDTFESMPREGRYTTGASCATAPVNY 299  
 QY 304 LSTVGSCTTLCPLHNOEVTADGTORCEKSKPCARCYGIGMOYIYANSKFTIGTELE 363  
 DB 300 VVTHGSCVNRACGPDYEV--BEDGIRCKCKDGRPKCKNGIGIGER--DTLSINATYIK 357  
 QY 364 -PACCKKIFGSLAPESFDDPASNAPALQPEQLQVEFLEITGYLYISAWPDSLPDL 422  
 DB 358 HFKCTAISGDLHLPVAFKGDSEFRTPPLPRELEIKTYKEITGFLIIQAMPDNDTL 417  
 QY 423 SVFONLOVIRGRILHNGAVSLTLQIGISWLGRLSRLSGSLALHINHTLCPVHTYPM 482  
 DB 418 HAFENLEIRKTYQHGFSLAVVGLNLTSLGLSLKEISQGDVITISGRRLCYANTINW 477  
 QY 483 DDLFRFNFTVSMILRVKVSASHLE--ECVGEGLACHOLCARGCHWGPGPTQCNVCSQFL 541  
 DB 478 KKLFGTNN-----QKTKIMNNRAEKCKAVNNHYCNPLCSGEGCGEPDCTSCQVVS 530  
 QY 542 RGDCVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGRPADQVACAHYKDPF 601  
 DB 531 RGRECEKCNILBEPRFVNSSECIOCHPECLPQAMNITCTGRGPDNCICAHYIDGP 590  
 QY 602 CVARCPGSGVKPDLGYMPIMKFPDEGACQPCPINCTHSQVLDLCKGCAEGRASPLNIV 661  
 DB 591 CVKTCGPMGIBENNTL--VMKTDANNVCHLHANCTTGACGPGLOGCEVWDSGKPIBIA 649  
 QY 662 SAVVGILLVVLGVVFGI--LIKRRQOKIRKTYMRRLLOETELVEPLTPSGAMPNQAQRI 720  
 DB 650 TGIIVGLLFIV--VALGIGLFMRBRHLYRKTLLRLLOERLVEPLTPSGAPQAQRI 708  
 QY 721 LKTELARKVVLGSGAGFTYKGIWIPDGEVANKIPVAIKVLRNTSPRANKELIDEAYVM 780  
 DB 709 LKTELARKVVLGSGAGFTYKGIWIPDGEVANKIPVAIKVLRNTSPRANKELIDEAYVM 768  
 QY 781 AGVGSFVYSRLIGICTSTVOLVQLMPYGLDHPVENRGLSDQLMNCQIANGMS 840  
 DB 769 ASVDNPHVCHRLIGICTSTVOLVQLMPYGLDHPVENRGLSDQLMNCQIANGMS 828  
 QY 841 YLEEDVRLVARDLAARNVLYVSPNHVKTIDFGLARLDIDETEVHADGKVPYIKMALESI 900  
 DB 829 YLEEDVRLVARDLAARNVLYVSPNHVKTIDFGLARLDIDETEVHADGKVPYIKMALESI 888  
 QY 901 LRRRTHQSDVMSGVTVMELMTGAPYDGIIPAREIPDLLEKERLPORPCTIDVYMI 960

DB 889 LHRITHQSDVMSGVTVMELMTGSKRPYDGIIPASDISILEKGERLPORPCTIDVYMI 948  
 QY 961 MYKCMIDSECRPRPRRELVSFSKMARDPQRFVYIQ--NEDLGPASPLDSTFYRLLED 1019  
 DB 949 MYKCMIDSECRPRPRRELVSFSKMARDPQRFVYIQ--NEDLGPASPLDSTFYRLLED 1008  
 QY 1020 MGDVLDAEYLVPOQGFPCPAPQAGGMVHRRSSSTRSGGDLTLGLEPSEEARPS 1079  
 DB 1009 MEDVVDADYELTPQGGF-----NSPST-----SRT 1034  
 QY 1080 PLASBEGAGSDVFDGLMGAKGLQSLPTHDPSPLGRSBDPTVPLPSET--DGYVAPL 1137  
 DB 1035 PLLSLSATSN-----NSVACINRNGSCRVVEDAFLOQYSSDPGCAVTEDNIDAF 1087  
 QY 1138 TCSPPQPEYVNPDPVPPSPRREGPLPAARFAGATLEAKTSLPGKGVKVDVFAFGAV 1197  
 DB 1088 ---FVPEYVNO--SVPKBPAGSVQNPVHNOPLHP-----APGRDLHYON--PHSNV 1133  
 QY 1198 ENPEYL--TPQGGAAPQHPAPAPAPDNLVYMDQ-----DP-----PERGAP 1240  
 DB 1134 GNPEYLVNTAQ-----PTCLSSGFNSPALMTQKSHQMSLDNPYQODFPFKETYPN 1184  
 QY 1241 STFKGTPAENPEYVGLDVP 1260  
 DB 1185 GTFKGP--PTAENAEYLRVAP 1203  
 RESULT 4  
 Q9YH40 PRELIMINARY; PRT; 1165 AA.  
 AC Q9YH40;  
 DT 01-MAY-1999 (Tremblrel. 10. Created)  
 DT 01-OCT-2000 (Tremblrel. 15. Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21. Last annotation update)  
 DE Receptor tyrosine kinase proto-oncogene.  
 GN XMRK.  
 OS Xiphophorus xiphidium.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorphia; Acanthopterygii; Percomorphia; Atherinomorphia;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 OX NCBI\_TaxID=8086;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RX MEDLINE=98241172; PubMed=9582016;  
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,  
 RA Altschimed J., Scharcl M.,  
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by  
 RT overexpression and mutational alterations.";  
 RL Oncogene 16:1681-1690(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RA Scharcl M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U534771; AAD10500.2; -.  
 DR HSSP; P11362; IRGK.  
 DR InterPro; IPR000345; CytC\_heme\_bind.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_Dkinase.  
 DR InterPro; IPR002174; Puriin-like.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TYRK; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN\_1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 2.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 KM ATP-binding; kinase; Transferase; Tyrosine-protein kinase.  
 SO SEQUENCE 1165 AA; 126614 MW; 7f7EE38D871A74E CRC64;

Query Match 39.7%; Score 2719; DB 13; Length 1165;  
 Best Local Similarity 45.4%; Pred. No. 4,5e-197;  
 Matches 582; Conservative 163; Mismatches 383; Indels 154; Gaps 31;

QY 1 MELALCMGILLALPPG-AAST-----QVCTGTMTKRLPASPETHLDMRLHYOGCV 55  
 DB 4 LLEL-----LILLLLISGRCCSTDPDRKVCQGTSTNQMTM--LDNHLKKMKMWSGCNV 56  
 QY 56 VQGNELTYLPTNASTFLQDIOEVGYVLAHNOVPLQRLIRVGTOLFEYNAL 115  
 DB 57 VLENLEITYTOBNQDLSLOSTIOEVGYVLAHNEVSTIPLVNLILIGNOLYEGNFTLL 116  
 QY 116 VLDNGDPLANTPTVYGASPGGLRELQRLSLTEILKGVLIQNPOLCYODTILMKDIFRK 175  
 DB 117 VMSNYQK-NPSSP--DYVQVGLKQLSLNLTFLISGVYVSHNPLLCNVERITNMWDIVDK 173  
 QY 176 NQOLALTLIDNRSACHPCSPMCKSGRCWGESSSDQSLRTVCAGGC-AACKGPLPD 234  
 DB 174 TSNPTMNLIPHAFERQCCQCDPGCVNCSWAPGPHCCFTLLCAEQCNRCRCRGPED 233  
 QY 235 CCHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPGRYTFGAS 294  
 DB 234 CCHCHAGGCTGPRATDCLACHDFDDGTCKDTCPPKITYDVSHQVVDNPIKITYFGAA 293  
 QY 295 CVTACPVNYLSTDVGCTPLVPLHNOEYTABDGTORCEKSPKARVCYGLM---QYI 350  
 DB 294 CVKCPSPNYVTE--GACVRSACAGMLEVD--ENGRKSKCRPCDVCRCVDGIGISLSMTI 351  
 QY 351 KANSKFIITELERFAGCKKIFGSLAFLESPFDGDASTAPLOPQLOVFEILEITGYL 410  
 DB 352 AVNSTNIG----SFSNCKTNGINDIILNRNSEFGDHYKIGPMDPHLMNLTIVKKEITGYL 407  
 QY 411 YISAMPDLSPLSVQNTLOVIRGRILHNGAYS--LTLOGISLMLGLRSRLSELGSLALIH 469  
 DB 408 VIMMPEMNTSLSVQNLLEIRGRITTFSRGFSFVVVQVSHLOMLLSLKEVSAGNVILK 467  
 QY 470 HNTHLCFVHTVPWDLFFRNFTVSPFWLVRPKVSASHLECEVGEGLACHOLCARGHCWP 529  
 DB 468 NTPQLRVASTINMRLFRSEDSQIEYDART-----ENQTCNNESESDGCWGP 514  
 QY 530 GPTQCVNCSQFLRGECEYBEGRVLOGLREYVNAHCLPCHECQCPQNGSVTCFEPZADQ 589  
 DB 515 GPTMCVSCLVHVRGRCVASCNLLQGBREDAVDGRCVQCHQECLVQTDLSLTCYGPGRAN 574  
 QY 590 CVACAHYDPPFCVARGSPGVKPDLSYMPIMKFPDEEGACOPCPINCHSCVVDLDKQCP 649  
 DB 575 CSKCAHFDGQPCIRCPHGMIGDDTL-IMKYLADKMGCCOPCHQNTQCGSGPPLSGCR 633  
 QY 650 AEORASPLTIVSAVVGILLVVLGVFGLIKRQOKIKRYTMRLLQETELVEPLTPS 709  
 DB 634 GD-IVSHSSLAVALSGLLITVIVALLIVLLRRRIK-RKTIIRLLQEKELVPLTPS 691  
 QY 710 GAMPQAOQRIKTELKRYKVLGSGAGFTYKGIWIDGDEVNKPVALIKVIRENTSPKA 769  
 DB 692 GOAPQOAFRLIKETEFKDRRLGAGAGTYVKGIMNDGENIRIPVALIKVIRENTSPKV 751  
 QY 770 NKEILDEAYVWAGVSPVYSRLGICLTSTVOQLVQOLMPYGLLDHVENRGRILSODIL 829  
 DB 752 NOEVDEAYVWASVDHPCRLGLICLSAVQVLQOLMPYGLLDHVENRGRILSODIL 811  
 QY 830 NMCQIAKMSYLEVDVLRVHDLAARNYLKVSNNHVKITDFGLARLIDETEHADGK 889  
 DB 812 NMCQIAKMSYLEVDVLRVHDLAARNYLKVSNNHVKITDFGLARLIDETEHADGK 871  
 QY 890 VPIKMALESTLRKFTHQSDVMSYGVTWELMTGAPYDGIIPAREIPDLLEKERLPQ 949  
 DB 872 VPIKMALESTLRKFTHQSDVMSYGVTWELMTGSKRPYDGIIPAREIPDLLEKERLPQ 931

QY 950 PPICTIDVYIMVWCMKIDSECRPRFRLVSEFSFMRARDPQRFVYIIONEDLGPASPLDST 1009  
 DB 932 PPICTIDVYIMVWCMKIDSECRPRFRLVSEFSFMRARDPQRFVYIIONEDLGPASPLDST 988  
 QY 1010 FRSLEDDDDMGDLVDAEYVLPQGFPCPDPAAGCMVHHRHSSSTRSGGDLTLGL 1069  
 DB 989 LFRSLSSDD--DVVDADDEYLL-----RYKRIN-RQGS----- 1018  
 QY 1070 EPSEEARSPVLAPEGSGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPV-PLPS 1128  
 DB 1019 -----EPCIPENGH-----PVRENSTALRTISDPONALEK 1049  
 QY 1129 ETDGVAPLTCSPQPEYVNOQDVRQP-----PSPR-----GPLP-AARPAGATLER 1175  
 DB 1050 DLDGH-----EYVNOQGSSTSRSLSDIYNVNYEDLTDGKCPVLSLSQEAETNFSR 1099  
 QY 1176 AKTISPGNGVAVQVFAFGAVENDEYLTPOGGAAPQHPHPAPSPAFDNLVYWDOPPE 1235  
 DB 1100 PEYLTNONSLS---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQ 1134  
 QY 1236 RGAPSTFKGTPTAENPEYGL 1257  
 DB 1135 TGAITGNGMFLPAAENLEYGL 1156

## RESULT 5

Q9W6F6 PRELIMINARY; PRT; 1137 AA.  
 AC Q9W6F6;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase (Fragment).  
 GN ERB4.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_Taxid=9031;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDRAIN;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.; AAD31764.1; -.  
 RT "Distribution of neurogranin-1 (nrg1) and erbb4 transcripts in  
 RT embryonic chick hindbrain."  
 RL Mol. Cell. Neurosci. 13:237-258(1999).  
 DR EMBL; AF121963; AAD31764.1; -.  
 DR HSSP; P11362; 1FGK.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001368; TNFR\_c6.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 1.  
 DR Pfam; PF02757; YLP; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 KM PROSITE; PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
 KW kinase; Tyrosine-protein kinase.  
 FT NON TER 1  
 SO SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC94F CRC64;

Query Match 38.9%; Score 2662.5; DB 13; Length 1137;

Best Local Similarity 46.3%; Pred. No. 8.4e-193;  
Matches 533; Conservative 171; Mismatches 352; Indels 95; Gaps 28;

```

Qy 161 LCYODTIWKDI FHKNOALALITDITNSRACHPSPCKSRCKGSESDCQSLTRVC 220
Db 3 LCFADTIHMODIVRNPMASNFTLVPTNGSSGGRCHKSCGTG-RCMGPTENHQQTLTKTV 61
Qy 221 AGGC-ARCKGLPTDCCHECCAGCTGPKHSDCLALPHNHSIGIELHCPALVTYNTDTF 279
Db 62 AEQCDGRCTGPGYVSDCHRECGCGSGRPDIDCFKMFNSGACVTCQPGTFVNPPTF 121
Qy 280 ESMNPBGRYTFGASCVTACPNYLSTDVSGCTLVCPHNOEVTADGTORCEKSKPCA 339
Db 122 QLEHNNAKYTYGAFCAVCKCPHNVF-YDSSCVACPSKMEV-BENGIMKCKPTDIDCP 179
Qy 340 RVCYGLGMOYIKANSKFIGITELE-FAGCKKIFGSLALPESFDDPASNAPLPEQLO 398
Db 180 KACGIGTGSU-VSAQOTVDSSNIDKFINCTKINGMLFLVYGIDHPYHTIAINPEKLN 238
Qy 399 VFETLEETGYLYISAMPDSLPLDSVFQNIYIRGRIHNGAYSLTLQIGISWGLASL 458
Db 239 IFQVIREITGYLNIOSPEENMTDFVFSNLVTIGRALYSGSLILKQOQITSLQFSL 298
Qy 459 RELSGALIHNTHLCEVHTVPMQDLRFNNFTVSFWLRVPKVA-----SHLECYGE 513
Db 299 KQISAGNIYITDNSLCYHTVMTSLF-----STPSQKTVIHRNKKXENCTAD 347
Qy 514 GLACHOLCARHGCWGPGTQCVNGSQPIRGQECVECKVLOGLPREYVNAHCLCPHPC 573
Db 348 GMVNCNELSSDGCWGPDOGLSCRFIRGRICIESCLYLQGEFREFANGSVCMEDCQC 407
Qy 574 QP-ONGSVTCFPGREADOCVACAHYKDPFCVACRPGVYKPDLSYMPKPFPEBEGACPC 632
Db 408 EKEMDNMTTCGPGEDHCTGCFHFDGPNCEKCEKCDGLQGANF--IKRIADEDECHPC 465
Qy 633 PINTHSCVDLDDKCC-----PAEQRASPLTISVAVV-GILLVVVLGVVFGIL 680
Db 466 HPNCTGCGRASHSDCIYPTWSTPLPQHAR-TPL--IAAGVIGLPIIVIMGLTFVAV 522
Qy 681 IKRQOQKIRKXTMRLLLOETELVEVLTPSGAMPNOAQRIKLETBLKRVKVLGSAFGTV 740
Db 523 VRRKSIK-KKRALRFL-ETELVEVLTPSGTAPNOAQRIKLETBLKRVKVLGSAFGTV 580
Qy 741 YKGIWIPDGENVYKIPVAILKVLRENTSPKANKIILDEAVYAVGVSPVSRLLIGLSTV 800
Db 581 YKGIWIPDGENVYKIPVAILKVLRENTSPKANKIILDEAVYAVGVSPVSRLLIGLSTV 800
Qy 861 QLVTLMPYGLLDHVRNKGRLSGQDLLNMCQIAKMSYLEDVRLVHRDLAANVLVK 860
Db 641 QLVTLMPYGLLDHVRNKGRLSGQDLLNMCQIAKMSYLEDVRLVHRDLAANVLVK 860
Qy 861 SPNVKVTDFGLARLLEDEKEYNADGKMPKMMALCIIHYRKTFHSDVMSYVITME 920
Db 701 SPNVKVTDFGLARLLEDEKEYNADGKMPKMMALCIIHYRKTFHSDVMSYVITME 920
Qy 921 LMTGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVVMIMVKKMIMSECRPRELVS 980
Db 761 LMTGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVVMIMVKKMIMSECRPRELVS 980
Qy 981 EFSMARDPORFVVIQNEE-LGPASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFPCP 1039
Db 821 EFSMARDPORFVVIQNEE-LGPASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFPCP 1039
Qy 1040 DRAGAGGMVHNRHSSTRSGGDLTLGLERSEEARSS--PLAP-SEGASDVFDGDL 1096
Db 880 PRYTSKRRIDSNRNOFYRQGYAAEGGV-FMPYRAGCIIPEAPVQAQTAELFEDTC 938
Qy 1097 GMAKAGLQSLPTDPSPLQRYSEDPVPLPS-----ETDGPALPTCSAPOPYYVNP 1149
Db 939 CNGTLRQVATLAKEDSDSTQYSAIDPTVFIPEVIRVGLDLBDGVTTPRKDKTDYLPV 998
Qy 1150 DVRRQPSRREGRLPA-RPAGATLTERAKTSLSPGNGVAVKVF-----AFGGAENP 1200

```

Db 999 EENPFVSRKXQDLOAVDNPEYHN-----APNGCPRAEDEYVNEPLVNTFANTLENA 1051

```

Qy 1201 EYLTPQGAAPQPPRPPAFNPNLYWDDPPERGA--PPSTKGRPT----- 1248
Db 1052 EYL-----KNMPEKAKKAFNDPDMYHNSLPPRSSTLOHDPDYLQEVSTKYFYXONGRI 1103
Qy 1249 -----AENPEYL 1255
Db 1104 RPIVANEPEYL 1114

```

RESULT 6  
P79754 PRELIMINARY; PRT; 1328 AA.

```

ID P79754
AC P79754
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erdb3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_Taxid=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
   rubripes";
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; Fu; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;

```

Query Match 33.3%; Score 2280; DB 13; Length 1328;  
Best Local Similarity 39.9%; Pred. No. 1.1e-163;  
Matches 516; Conservative 162; Mismatches 417; Indels 198; Gaps 33;

```

Qy 9 WGLLLALLRP--GAASDQ-----VCTGDMKRLRPASPEPTLMDLHLYOGCOVQGNLEL 62
Db 4 WRLILMCVARSRLRAASSSTQGAVCEGTONGISSTGSGQENQYNLNDKRYKGGIINGNLEI 63
Qy 63 TYLPTNLSLFLQDIOEVQGYVLIAHNOVRQPLQRLRIYVGTQLFEDNYALVALVNDGPP 122
Db 64 TQISNDMDFSLKTRIREVTGTVLLIAMHNFQEIPLGLAVINGNSLYERFALSVFLN--- 120
Qy 123 LANTTPVTGASPGGLRELQSLSLTEILKGVLIQRPOLCYQDITLWKDIFHKNOALALT 182
Db 121 ----YPKG--PSGILNQGLNLTLEILDGVOIINNKRLRYGPMWYWRDII-RNNDADIE 173
Qy 183 LIDITNSRACHPSPCKSRCKGSESDCQSLTRTVAGGC-ARCKGLPTDCCHECA 241
Db 174 IQNGERGVCV---KSC-GNVCWGRKQCCQILTKTVACAPQCNDRCFSTSPDCCHISCA 229
Qy 242 AGCTGPKHSDCLALHFNHSGICELACPALVTYNTDFESMPNPEGRTYFGASCVTACPY 301
Db 230 AGCKGRLDIDCFALRLFNDSGACVPGCPQTLIYNKQTFQWETNPNPAKYQSGISCVSQPT 289

```



```
QY 302 NYLSTDVSCCTLVCPNLHNOEYTAEDGTOR-CEKSKPCARVCYGLGMQYITANSKFTIT 360
DB 290 HFV-VDSSSCVSVCPDMEV--ERGSQOCELCSGCPKVCCEGTAE--QRTVDSS 342
QY 361 ELE-FACCKIIFGSLAFLPESFDGDPASNTAPLQEOLOVETLEIITGYLYISAMPDL 419
DB 343 NIDFINCTKIQSSLHFLVTGILGDDFKQVPEPLDAKKLEVRITREITDILNQSMPLEL 402
QY 420 PDLSPFQNLQVIRGRILHNGAVSYLTQGLGISWGLRSRLBSGLALIHNTHLCFVHT 479
DB 403 NDLVSFSSLTITIGRSLEKRFSLMWMRIPTLISLGRSLREISDSQSVYISQNAHLCHYHT 462
QY 480 VPMQDLRFNNFTYSPMLRVPKVASH-LEECVGBGLACHOLCARGHWGCGPPQCVNCS 538
DB 463 VNMQLFRGSR-----VRANSLSNRMACEVADGRVCDPLCSDSGCMGCGPPQCLISCR 516
QY 539 QFLGQSCVEECRVLOGLPREYVNAARH-CLPCHPECPQONSVCYCFEADQCVACANHYK 597
DB 517 NYSHGCTVAGCHNSGIPREFAGLVGCVACHPECKQTQKASCTGFAGDECMACIKYFR 576
QY 598 DPPECVACPSGVKPDLSYMPIMKFPDEGACOPCPINCTHSQVLDLDDKGPAPORASPL 657
DB 577 DGPYCMSSCPAGVA-DGEKGLIFKFPNREGHCEPCHQCTQCGSGPLNDC--LEAARL 632
QY 658 TSIVSAVVGILLVVLGVF-----GILKROQKRTKRTMRLLQETLVEPLTSG 710
DB 633 TISSGQTGIALGVPAGLIFCLVLPFLGMLYHRGLAIRKKAMRYLESSESFBPLGP-G 691
QY 711 AMPQAOAMRIKETELRKVKVYLSGAGFTVYKGIWIPDGEVVKI.PVAIKVLRNTPSPAN 770
DB 692 EKGKVARILKPSDLKIKPLGSGVGTGSKGFWIPGSEVTKIPVALKITIQDSSGRQTF 751
QY 771 KEILDEAYVAVAGVSPYVSLGLICLSTVQLVQLMPYGCGLDHVRENRGLSQDLN 830
DB 752 TEITDHLISMOSLDHPYIVRLGICPQTCQLVQLTQSLSHGSLLEHIRQKHSIDLPQRLN 811
QY 831 WCMQIAGMSYLEVYRLVHRDLAARNLVYKSPNHYKITDPELALILIDETETIADGKV 890
DB 812 WCVQIAGMYLLEHRYVHKMLAARNILKNQYQVQISDGVADLLVPDDKYYVSEKTK 871
QY 891 PIKMALESILRRRFTHOSDVSYGVTWELMTFGAKPYDGI.PAREI.PDLLEKGERLPOP 950
DB 872 PIKMALESILFRKRYTHOSDVSYGVTWEMMSFGAEYASVQOEVPVLEKGERLSQP 931
QY 951 PICTIDVYIMVWKCMWIDSECRPRFRELVSFSSPMADPQRFVJIONEDLAPASPLDSTF 1010
DB 932 AICTIDVYVWVKCMWIDENIRPTFEKLASDFTRMADPRVYLVIRMEG----- 980
QY 1011 YRSLLEDDMDGLVDAEEYLVPOQGFPCPPAPAGCMVHHRRSSSTRSGGDLTLGLE 1070
DB 981 -----EDSGMGEFL-----RRGSR-----GILLEADLE 1003
QY 1071 PSEEARSPPLAPSEAGSDVFDGDLGMG---AAKGIQSLPTHPSPLQ-----R 1117
DB 1004 EDEBE-----GIGDRFAITPSLQSPSSTWSSTBSQINSYVMTQLR 1042
QY 1118 YSEDPVPLPSETDGYVAPLTCSPQ- EYVNO-----PDVPRQPPSPREP 1162
DB 1043 YD-----FAVSGGHHGYLPMSPSPYDITRQLWYQSRSLSSVYTLDRSFRSSSEAE 1096
QY 1163 L--PAARPAATLERATTLSPGKNVYKDVPAFGCAVENPEYLTPOGGAAPQHPAPAS 1220
DB 1097 LCEBDGACAGIFRVR-----FQSEKGN-----PQGS----- 1122
QY 1221 PAFDNLVYWDODPERGAPSTFKGTPTAENPE 1253
DB 1123 -----QQRKLSSTASSPSSFTMADEDE 1146
```

```
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SDA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ01655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Purin_like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.4%; Score 209.5; DB 5; Length 1433;
Best Local Similarity 32.4%; Pred.No. 3.9e-143;
Matches 473; Conservative 194; Mismatches 386; Indels 409; Gaps 39;

QY 26 CTGIDMLRLRPASFTLHDLRLHYQCGVYQSLTYLPTNMSLSLTDIOYQGYVL 85
DB 1 CIGNGMSVAPANEHYKMLRDRYNTCTYVDGNLEITWIONITDNLQHIRVETGYVL 60
QY 86 IAHNVQVPLQRLRIYRGTLF-----EDNYALAVLDNGDPLNNTTPTVGASPGGLREL 140
DB 61 ISLYDLPOVILPRLQITIRGRTTFKLNMKEEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGVLIQNPOLCYODITLWMDI-FHKNNQALTLITDNRSPACHPCSPMC 199
DB 105 ELPLRDLILGGSVGFNNYNLCHKMSINMEILILAPQSMQYTNFSSPREVCPCHSC 164
QY 200 KGSRCWSESSEDCSLRRTYACAGCA--RCKGRLPTDCHEQCAAGCTGPRHSCLACLH 257
DB 165 EVG-CWGEAGNCCRFKLNCSPOCSQGRCPGRPRBCHLFCAGAGCTGTPQSDCLACKN 223
QY 258 FNHSGICELCPALVYNTDTPESMPNDEGYTGASCVYACAPYNYLSTDVGSCITVCP 317
DB 224 FYDDGVCKQECFPQOINPTNYFEPNPDGKIVAGATVAKCP-EHLIKDNGACVRKCPK 282
QY 318 HNOEVTAEDETQRCSEKSKPCARVCYGLGMQYITANSKFTIGITELFAGCKKIFGSLAFL 377
DB 283 GKMPQNSE-----CVPKGVCPKTCPEGI-----VHSDNIG-----NYKDCITIIIGSEIL 329
QY 378 PESTDGPASNT-----APLOPEOLOVETLEIITGYLYISAMPDLSLVFQNLQ 429
DB 330 DQSFDFQOQVYTNFSGPRYIKIDPRLVFSFTVKEITGFINIOAHHPNFTLLNYFNNLE 389
```



```

RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC027080; AA027080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match
Best Local Similarity 88.0%; Pred. No. 1.7e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954
DB 1 MALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 60

QY 955 IDVMIWIKCMIMISECPREFELVSEFSRMAARDPQFVIVIONEDLGPASPLDSFFYSYL 1014
DB 61 IDVMIWIKCMIMISECPREFELVSEFSRMAARDPQFVIVIONEDLGPASPLDSFFYSYL 120

QY 1015 LEDDDMGDLVDAEEXLVLPQGFPCPDPAFGAGVNHHRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDAEEXLVLPQGFPCPDPAFGAGVNHHRSSSTRSGGDLTLGLEPSEE 180

QY 1075 EAPRSPLAPSEAGSDVFDGLGKGAAGLQSLPTHDSPLQRYSEDPVLPSPSTGGYV 1134
DB 181 EAPRSPLAPSEAGSDVFDGLGKGAAGLQSLPTHDSPLQRYSEDPVLPSPSTGGYV 240

QY 1135 APLFCSPOPEYVNOPODVARPOPSPRECPPLPARPAGATLERAKTSLPKNGVVKOVFAFG 1194
DB 241 APLFCSPOPEYVNOPODVARPOPSPRECPPLPARPAGATLERAKTSLPKNGVVKOVFAFG 300

QY 1195 GAVENPEYLTPOGGAAPQHPHPAPFSPAFDNLVYWDQDPERKAPSTFTKGTPTAENDEY 1254
DB 301 GAVENPEYLTPOGGAAPQHPHPAPFSPAFDNLVYWDQDPERKAPSTFTKGTPTAENDEY 360

QY 1255 LGLDVVPV 1261
DB 361 LGLDVVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
ID Q86712;
AC Q86712;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Polypeptin.
GN POLYPEPTIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL: S69372; AAC60725.1; -.
DR HSSP: P03322; 1A6S.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR004026; Retro_M.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

```

```

DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914FE1D63 CRC64;

Query Match
Best Local Similarity 54.8%; Pred. No. 1.3e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 575 PONGSVTCFGEADOCVACAHKYDPPFCVAPRSGVKPPLSYMPILKFPDEGACOPCP 634
DB 141 PEETPTPTGTP--DHCMCAHFIDPCHVCYACPAVLGENDTL-VKVIADANAVCOLCHP 197

QY 635 NCTHSCVDLDDKGCPEAPRASPPLTISVAVV-GILLVVLGVVFGILLKROOKIRKTYM 693
DB 198 NCTHSCVPLGEGCP---NGSKTSPISAGVVGGLCLLVVGGIGLGYLARR-HYVKRTL 253

QY 694 RRLDQETLVEPLTSGAMPNOQMRILKETELRKVKYLGSGAFGTVYKGINIPGSENK 753
DB 254 RRLDQETLVEPLTSGAMPNOQMRILKETELRKVKYLGSGAFGTVYKGINIPGSENK 313

QY 754 IPVAIKVRENTSPKANKELIDEAYVMAGVSPYVRLIGLCTSTVOLVTLQMPYGLL 813
DB 314 IPVAIKELRENTSPKANKELIDEAYVMASVDNPRVCRLLGLCTSTVOLVTLQMPYGLL 373

QY 814 DHVRENRLGSODLLNWCMTAKGMSYLEDVRLVHRLDAARNVLKSPNHVKITDFGLA 873
DB 374 DYIREHKONISQOYLNNVCQAKGMNLEBRVLVHRLDAARNVLKTPQHKITDFGLA 433

QY 874 RLIDIDETEHYADGKVPILKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGI 933
DB 434 RLIDADEKEHYAEGKVPILKMALESILRRRTTHOSDWSYGVTVWELMTFGSKPYDGI 493

QY 934 AREIPDLLEKBERLPQPICTIDVMIWIKCMIMISECPREFELVSEFSRMAARDPQFV 993
DB 494 ASEISSVLEKBERLPQPICTIDVMIWIKCMIMIDADRPRKRELIAERSKMARPPRL 553

QY 994 VIQ-NEDLGPASPLDSTFYRSLLEDMDGDLVDAEEXLVLPQGFPCPDPAFGAGVNHHR 1052
DB 554 VIQDERKHLRPSPTDSKYFRTLMEEDMEDIDVDAEEXLVLPQGF----- 598

QY 1053 HRSSSTRSGGDLTLGLEPSEEARSPPL-----APSEAGSDVFDGLGKGAAGLQSL 1107
DB 599 -NSPST-----SRTPLSLSLATSNNSNATNCID-----RNGQGH 631

QY 1108 PTHDPSPLQRYSEDPVLPSPET--DGVAAPLTCSPOPEYVNOPODVARPOPSPRECPRLA 1165
DB 632 PVREDSFYQRYSSDPTGNFLBESIDDFL-----PABEYVNO-LMPKPS----- 675

QY 1166 ARPAGATLERAKTSLPKNGVVKOVF-----AFGAVENPEYL 1203
DB 676 -----TAMVONQIYNNISLTAISKLPRDSRYQNSHSTAVIDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
ID Q86714;
AC Q86714;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";

```

RL Oncogene 9.1307-1320(1994).  
 DR EMBL; S69372; AAC60727.1; -.  
 DR HSSP; P11362; 1FGK.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase.1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase.1.  
 DR SMART; SM00219; TyrcK.1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP.1.  
 DR PROSITE; PS0011; PROTEIN KINASE\_DOM.1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR.1.  
 DR Tyrosine-protein kinase.  
 DR NON TER  
 FT 1  
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;  
 Best Local Similarity 55.4%; Pred. No. 12e-121;  
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQVCACAHYKDPFCVACPSGVKPDLSYPMFWPDEGACOPCPINCTHSCVDL 643  
 DB 1 GP--DHCKKCAHFIDGPHCVKACPAVIGENDTL-VWKYADANAVCQLCHPCTGCKGP 57  
 QY 644 DDKGCPAEQASPLTSIVSAVV-GILLVVLGVFEGLIKRQOKIRKRYMRLLQETEL 702  
 DB 58 GLEGCP--NSKSTPSIAGVGVGLCLVWVGILGILYLR-R-HIVRKRTLRLLQEREL 113  
 QY 703 VEPLTPSGAMPNOAMRIKTELAKVYLGSGAGTYKGIWIDGENVKIPVAKYLR 762  
 DB 114 VEPLTPSGEAPQAHRLIKETEFKKVKVLSGAGTYKGIWIDGENVKIPVAKYLR 173  
 QY 763 ENTSPKANKEILDEAVYVAGVSPYVSRLLGICLTSTVQLVQLMPYGLDHWENGR 822  
 DB 174 EATSPKANKEILDEAVYVAGVSPYVSRLLGICLTSTVQLVQLMPYGLDHWENGR 233  
 QY 823 LGSQDLMLWCMQIAGMSYLEDVRLVHRDLAARNLVKSPNNVKTITDGLARLDIDETE 882  
 DB 234 IGSQDLMLWCMQIAGMSYLEDVRLVHRDLAARNLVKSPNNVKTITDGLARLDIDETE 293  
 QY 883 YHAGQKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAPYDGIIPARELPDLE 942  
 DB 294 YHAGQKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAPYDGIIPARELPDLE 353  
 QY 943 KGERLPQPICTIDVYIMVCMIDSECRPRELVESEFSRMARDPQRFVYIQ-NEDIG 1001  
 DB 354 KGERLPQPICTIDVYIMVCMIDSECRPRELVESEFSRMARDPQRFVYIQ-NEDIG 413  
 QY 1002 PASPLDSTFTYRSLLEDMDGDLVDAEYLVPOQGFCCDPAPAGAGMWHHRSSSTRSG 1061  
 DB 414 LPSEPTDSKFYRSLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454  
 QY 1062 GGDLTGLGPESEEEAPRSPPL-----APSEGAGSDVFPDDLGMAKGLQSLPTHDPSPLO 1116  
 DB 455 -----SRTPLLSSLSLATSNNKATKID-----RNGQGHVVRDSTFVQ 491  
 QY 1117 RYSEDPVPLPSET--DGVAAPLTCSPQPEYVNOQDVAPQPPSPREGPLPARPAGATLE 1174  
 DB 492 RYSEDPVPLPSET--DGVAAPLTCSPQPEYVNOQDVAPQPPSPREGPLPARPAGATLE 1174  
 QY 1175 RAKTLSPGKNGVADV-----AFGAVENPEYL 1203  
 DB 527 -----TAMVQNIYNNISLTALSKLPMSRYONSHSTAVDNPEYL 566

RESULT 12

QY 08WYVO PRELIMINARY; PRT; 412 AA.  
 AC 08WYVO;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Hypothetical 44.7 kDa protein.

GN PP3659.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OK NCBI\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
 RA Wan D.F., Gu J.R.;  
 RT "Novel human cDNA clones with function of inhibiting cancer cell  
 RT growth."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF318349; AAL5856.1; -.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00069; pkinase.1.  
 DR Pfam; PF02757; YLP.2.  
 DR ProDom; PD000001; Euk\_pkinase.1.  
 DR SMART; SM00219; TyrcK.1.  
 DR PROSITE; PS00108; EF\_HAND; UNKNOWN.1.  
 DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM.1.  
 DR Hypothetical protein.  
 KW SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;  
 Best Local Similarity 80.5%; Pred. No. 2.8e-120;  
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPARELPDLEKGERLPQPICT 954  
 DB 1 MALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPARELPDLEKGERLPQPICT 60  
 QY 955 IDVIMVCMIDSECRPRELVESEFSRMARDPQRFVYIQ-NEDIGPASPLDSTFTYRSL 1014  
 DB 61 IDVIMVCMIDSECRPRELVESEFSRMARDPQRFVYIQ-NEDIGPASPLDSTFTYRSL 120  
 QY 1015 LEDDDMDGLVAEYLVPOQGFCCDPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEE 1074  
 DB 121 LEDDDMDGLVAEYLVPOQGFCCDPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEE 180  
 QY 1075 EAPSPPLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGV 1134  
 DB 181 EAPSPPLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGV 240  
 QY 1135 APLTCSQPEYVNOQDVAPQPPSPREGPLPARPAGATLEPAKTLSPGKNGVADVAFG 1194  
 DB 241 APLTCSQPEYVNOQDVAPQPPSPREGPLPARPAGATLEPAKTLSPGKNGVADVAFG 300  
 QY 1195 GAVENPEYLTQGGAAAPQ-----HPPA---FSPAFLNL 1226  
 DB 301 GAVENPEYLTQGGAAAPQ-----HPPA---FSPAFLNL 1226  
 QY 1227 YVMD-ODPPER-----GAPSTFKGTPTAEN 1251  
 DB 361 YVMD-ODPPER-----GAPSTFKGTPTAEN 1251

RESULT 13

QY 064895 PRELIMINARY; PRT; 962 AA.  
 AC 064895;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Gag, v-erb-A, v-erb-B protein.  
 GN GAG, V-ERB-A, V-ERB-B.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 OK NCBI\_Taxid=11861;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RX MEDLINE=90206603; Pubmed=1969616;  
 RA Bruskin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;  
 RT "Six amino acids from the retroviral gene gag greatly enhance the  
 RT transforming potential of the oncogene v-erb-B.";  
 RL Oncogene 5:15-24(1990) .  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY) .  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL; X52209; CA36459.1; -  
 DR EMBL; X52211; CA36459.1; JOINED.  
 DR HSSP; P10828; ZNLF.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR000536; Hormone\_rec\_1ig.  
 DR InterPro; IPR001723; Stehhm\_receptor.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR InterPro; IPR001628; Znf\_Casteroid.  
 DR Pfam; PF00104; hormone\_rec\_1.  
 DR Pfam; PF00069; pkinase\_1.  
 DR Pfam; PF00105; zf-C4\_1.  
 DR PRINTS; PRO0398; STRDHORMONER.  
 DR PRINTS; PRO0047; STROIDFINGER.  
 DR ProDom; PD000001; Euk\_kinase\_1.  
 DR ProDom; PD000035; Znf\_Casteroid; 1.  
 DR SMART; SM00430; HOL1\_1.  
 DR SMART; SM00219; TykC; 1.  
 DR SMART; SM00399; Znf\_C4\_1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; DNA-binding; Nuclear\_kinase; Receptor;  
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;  
 KW Zinc-finger.  
 QO SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B95CE CRC64;

	Query Match	24.2%	Score 1653.5	DB 15	Length 962
	Best Local Similarity	51.7%	Pred. No. 2.1e-116		
	Matches 358	Conservative	73	Mismatches 142	Indels 119
				Gaps	18
Qy	547	VEECRVQLGGLPRE-VYVNAH-HCLP-----CHPQC	574		
Db	354	IEKQGESILLAFHYINVRKINIHFWSKLMTKADLRMIGAYHARLRLLHMKVECPITLS	413		
Qy	575	PONGSVTCFGEADQCAAHYKDPFCVCARCPSGKPDLSYMPIKPEDEGACOPCI	634		
Db	414	POE-----VGB--DHCMCAHFIDGPHCHKCAPAGVLENDTL-VMKYADANAACQLCHP	465		
Qy	635	NCHTSYVDLDKCGPAEGRASPLTSYSAV-GLLVVVLGVFSLIKRQOKIRKTYM	693		
Db	466	NCTRGCGKPGLEGCP--NGSKPTSIAGVVGGLCTLVVGLIGLTYLRH-HIVRKRTL	521		
Qy	694	RRLQETELVEPLTPSGAMPNQAOMRIKETELRKVKYLGSGAGFTVYKGIWIPDGENVK	753		
Db	522	RRLQGEELVEPLTPSGEAPVQAHRILIKETEPKKVKYLGSGAGFTVYKGIWIPDGENVK	581		
Qy	754	IPVAIKVLRENTSBRANKELIDEAYVMAVGSPPVYSRLGLCTLSTVOLVQLMPYGCIL	813		
Db	582	IPVAIKELREATSPRANKELIDEAYVMAVSDVNPVYCRLLGLCTLSTVOLVQLMPYGCIL	641		
Qy	814	DYHRENGRGLGSQDLLMWCIOIANGSYLEDVRLVHRDLAARNVLYKSPNHVKTIDFGLA	873		
Db	642	DYIREHNDIGNSQYLLMWCIOIANGMYLLEERHVAHRDLAARNVLYKTPQVKTIDFGLA	701		
Qy	874	RLLDIDETEHADGKVPKIMMALESILRRFTHOSDWSGTVVWELMTGCAAPYDGP	933		
Db	702	KQLGADKEKTHAEGKVPKIMMALESILHRIYTHOSDWSGTVVWELMTGSKRYPGIR	761		
Qy	934	AREIPLDLKEGERLPORPCTIDYVMIMVCMITDSECRPRPRELVSFSSMADDPORV	993		
Db	762	ASEISSVLEKGERLPORPCTIDYVMIMVCMMSGASDRPRFRELIAFSMAADDPRLY	821		
Qy	994	VIO-NEDLPASPLDSTFYFSLSEDDMDGLVDAAEYLVPOQGFCDPAPAGAMVHR	1052		
Db	822	VIOGDERMHLPSIPDSCFYFRTLMEEEMEDIVDADEYLVLPPOGFF-----	866		

```

QY      1053  HRSSTKSGGDDLTLGLEPSEEEKPRSPPLAPSEAGSDVDFPGDLGMGAAGKGLSLPHHPD  1112
          |||
Db      867   -NSPST-----SRPPLSLSLATSN-----NSATKCIDNNGCH--  898
          |||

QY      1113  SPLQRYSDDTPVLPSETDGVVAFLTCSQPEEYVNOQDVNPQPSPREGCLPARAPGAT  1172
          |||
Db      899   -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVNOQIYNYISLT  936
          |||

QY      1173  -LERAKTLSPGKNGVWVDVFAFGAVENPEYL  1203
          : : :
Db      937  AISKLPMDSDRYN-----SHSTAYVDNPEYL  961
          : : :

```

RESULT 14			
ID	085468	PRELIMINARY;	PRT; 545 AA.
AC	085468;		
DT	01-NOV-1996	(TREMblrel. 01, Created)	
DT	01-NOV-1996	(TREMblrel. 01, last sequence update)	
DT	01-MAR-2002	(TREMblrel. 20, last annotation update)	
DE	Avian Erythroblastosis virus (Tsj41) v-erbB gene.		
OS	Avian erythroblastosis virus.		
OC	Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.		
NC	NCBI_Taxid=11861;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=88217326; PubMed=2897102;		
RX	Scotling P., Vennstrom B., Jansen M., Graf T., Beug H., Hayman M.J.;		
RT	"Common site of mutation in the erbB gene of avian erythroblastosis		
RL	virus mutants that are temperature sensitive for transformation.";		
DR	Oncogene Res. 1:265-278(1987).		
DR	EMBL; X06943; CAA30024.1; -.		
DR	HSSP; P1362; 1FGK.		
DR	InterPro; IPR000719; Euk_pKinase.		
DR	InterPro; IPR001245; Tyr_pKinase.		
DR	Pfam; PF00069; pkinase; 1.		
DR	ProDom; PD000001; Euk_pKinase; 1.		
DR	SMART; SMO0219; TyrKc; 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.		
DR	ATP-binding; Transferase; Tyrosine-protein kinase.		
SO	SEQUENCE 545 AA; 60899 MW; 140DCE8CCAF8AF4 CRC64;		

Query Match	24.0%	Score 1645	DB 15	Length 545
Best Local Similarity	54.9%	Pred. No. 4e-116		
Matches 345	Conservative 10	Mismatch 121	Indels 92	Gaps 15
Qy	584	GREADQCVAAAHYDPPFCVARGCPSPGKLPDLSWPIKPFDEGACOPCINCHSCVDL	643	
Db	1	GP--DHCKMCAHPIDGHPCVKACBPAGVLGNDTL-VKRYADAAAVCOLCHPACTTRCGKP	57	
Qy	644	DDKCPAQRASPLSTIVSANV-GILLVVVLGVFGLIKRROOKIRKTYMRRLLOETEL	702	
Db	58	GLESCP---NGSKTPSIAGVGVGLCLTVVVGIGIGLYLRR-HVKKRTLRLLOREL	113	
Qy	703	VEPLTPSGAMPNOAMRIIKETELRKVYLSGSAFGVYVKGVIIPGGENVKIIVALKVL	762	
Db	114	VEPLTPSGEAPNOAHILKILETEFPKKVYKLVGFAGFYVKGWIMPEGETKTIIVALKELR	173	
Qy	763	ENTSPKANKEILDEAVYVMAGVSGSPVYSRLIGICLTSTVOLVLTQMLMPGCLLDVHRENRG	822	
Db	174	EATSPKANKEILDEAVYVMAVDNPHNCRLLGICLTSTVOLITQMLMPGCLLDVIRHKON	233	
Qy	823	LGSGODLLNMCQIAKGMSTYLEDVRLVHRDLAARNVLVKSBNHYKITDFGLARLLIDETE	882	
Db	234	IGSQYLLNMCVQIAKGMNYYLEENHLVHRDLAARNVLVKTPODYKITDFGLAKOLGADPEKE	293	
Qy	883	YHADGKGVPIKMMALESIILRRFTTHSDVWSGVVTWELMTFGAKRYDGIIPAREIPDLE	942	
Db	294	YHADGKGVPIKMMALESIILHRIYTHSDVWSGVVTWELMTFGOSKRYDGIIPASEISVYLE	353	

```

OY 943 KGERLPPICTIDVYIMVWCMIDSECRPRFRELVSFMSMARDPQRFVIO-NEDIG 1001
DB 354 KGERLPPICTIDVYIMVWCMIDSECRPRFRELVSFMSMARDPQRFVIO-NEDIG 413
OY 1002 PASPLDSTFFRSLDDDDMGDLVDVAEVYLPQOQFFCPDPAPAGAGMHHRRSSSTRSG 1061
DB 414 LPSPTDSKFYRLTMEEDMEDIVDADEVLPVPHQGF-----NSST--- 454
OY 1062 GDDLTLGLEPSEEEAPRSP-----APSEAGSDVFDGDLGMAKGLSLPTHPSPILQ 1116
DB 455 -----SRPDLSSLATSNKATNCIDRNG-----H----- 481
OY 1117 RYSEDPTVPLPSETDGYVAPLTCSPQEPYVNOPDVRPOPSPREGLPAPAPAGAT-TER 1175
DB 482 -----PYREDEGFL-----PAPEYVNO--LMPKKPSTAMVQIQYVYISLTATSK 523
OY 1176 AKTISPGKNGVKVAFAGAVENPEYL 1203
DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

```

## RESULT 15

O9WVFS PRELIMINARY; PRT: 655 AA.

AC O9WVFS: 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)

DE Epidermal growth factor receptor (Epidermal growth factor receptor

isoform 3).

GN EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LIVER;

RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,

RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RA "Alternative transcripts from the Human and Mouse EGFR Genes Encode

RT Carboxy-terminal truncated Receptors."

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RC STRAIN=C3H/101, 129/SVJ; AND 129/SVEYTAC.

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,

RA Sinclair C.S., Fearall R.S., Green P.J., Yee D., Lampland A.L.,

RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,

RA Maithle N.J.;

RT "Comparative genomic sequence analysis and isolation of human and

RT mouse alternative Egfr transcripts encoding truncated receptor

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LIVER;

MDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Toh M., Ishii Y.,

RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamashita I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,

RA Kuehl P., Lewis S., Marcuso Y., Nikaido I., Peesle G., Quackenbush J.,

RA Schirral L.M., Staab F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bulc C., Fletcher C., Fujita K., Gariboldi M.,

RA Guerinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Marzarelli U., Mombetters P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Wetz C., Whiteaker C., Wilmshurst L.,
RA Wymshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuke S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL, AF124513; AAD4149.1; -.
DR EMBL, AF275366; AAG28047.1; -.
DR EMBL, AF275365; AAG28047.1; JOINED.
DR EMBL, AF275365; AAG28047.1; JOINED.
DR EMBL, AK004944; BAB23688.1; -.
DR EMBL, AK004944; BAB23688.1; -.
DR EMBL, AK004911; BAB23662.1; -.
DR MGD: 95294; Egfr.
DR InterPro; IPR004944; EGFR_L_domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

Query Match 21.6%; Score 1475.5; DB 11; Length 655;  
Best Local Similarity 43.7%; Pred. No. 3.8e-103;  
Matches 283; Conservative 99; Mismatches 232; Indels 33; Gaps 11;

```

OY 11 LLLALPFGAA--STVCTGTDMLKRLPASPTLMDMLRHLYOGCOVVOGNETLYLPTN 68
DB 14 LLLALPFGAA--STVCTGTDMLKRLPASPTLMDMLRHLYOGCOVVOGNETLYLPTN 73
OY 69 ASLFLDIOEVGVYLLAHNOVROVPLORLRYVGTQLFEDNYLALVLDNDPLNNTTP 128
DB 74 YDLSLTKITQEVAGVYLLALNTVERKIPLENQIIRGNALYENTYALALISN----- 124
OY 129 VTGASPGGLRELQRLSTLEILKGVLIQRNPOLCYDPTILMKDI----FKKNQALTLI 184
DB 125 -YGTNRTELRELPRNLEIILGAVRFNNPILCMMDITQIRDIQVNFMSMSMDL--- 180
OY 185 DTNSRBRAPRSPSPCKSGRCMGESSEDQSLTRYVACAGCA-RCKGRLPTDCEHQCAAG 243
DB 181 -QSHRSSCPKCDPFCPNKSGWGGEGENCQKLTKIKIQAQCGHRCGRSPSDCHNQCAAG 239
OY 244 CTGPHSPDCLACLFHNSGICELHCPALVTYNTDFFESMPDEGRYTGASCVTACPNY 303
DB 240 CTGPHSPDCLACLFHNSGICELHCPALVTYNTDFFESMPDEGRYTGASCVTACPNY 299
OY 304 LSTDVGSCTVCPPLHNOEVTADGTORCEKSKPCARVYCGIMQYIRANSKFTIGITELE 363
DB 300 VVTDHSGCVRACGPDYEV--BEDGIRKCKKDDGPRKVCNKGIGEFK-DTLISATNIK 357
OY 364 -FACCKIKFGLALPSPFDGDPASNTAPLOPELOVETLEITGVLYIYISAMPDSDL 422
DB 358 HFXYCTAISGDLHILPAFKDSFTPTPRELEIKTYKEITGFLIIQAMPDNTDL 417
OY 423 SVFONLOVIRGRIHNGAVSLTLOGLISWGLSLRELSGLLIHNHTLCPFHNTVPW 482
DB 418 HAFENLEIRKTRTQHOQFSLAVGLNITSLGLSLKEISGDVYISGNRLCYANTINW 477
OY 483 DQLFRFNNFTVSWLRAVPKVSASHLE-BCVGEGLACHQLCARCGHMGPGPTQVCVCSQFL 541
DB 478 KKLFGTRN-----QTKIMNNNAEDCKAVNHVNCPLGSSBECGWGEPEDVCSQVNS 530
OY 542 RGQGVCECRVLOGLPREYVNAARCLFCHPCQVQNSVTCFGEADQVCAHYKDPF 601
DB 531 RGRCEVCKNLTLEBPPEFENSECICQHPCLPQANNTTCTGGRPDNCCIOCAHYIDGPH 590
OY 602 CVARCPGKVPKDLSTYMPIMKPRDEGACQCPRICTHSCVDLDKGC 648
DB 591 CVKTCPRAGINGENNTL-VKXADANNVCHLCHANCTTGACGPGLOGC 636

```

Search completed: July 22, 2003, 09:01:28  
Job time : 54.5887 secs

---







```
QY 781 AGVSPVYSRLIGICTSTVQLVQMPYGCILDHVRNRRGLSGQDILNMCQIAKMS 840
DB 775 AGVSPVYSRLIGICTSTVQLVQMPYGCILDHVRNRRGLSGQDILNMCQIAKMS 834
QY 841 YLEEDVRLVHRDLARNLVKS PNHVKITDFGLARLIDIEVEYHADGKVPDKMMALESI 900
DB 835 YLEEDVRLVHRDLARNLVKS PNHVKITDFGLARLIDIEVEYHADGKVPDKMMALESI 894
QY 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKEGRLLPQPICTIDVMI 960
DB 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKEGRLLPQPICTIDVMI 954
QY 961 MYKWMIDSECRPFRELVSEFSKMRADPQRFVVIQNEIDLCPASPLDSTFYRSLIEDDM 1020
DB 955 MYKWMIDSECRPFRELVSEFSKMRADPQRFVVIQNEIDLCPASPLDSTFYRSLIEDDM 1014
QY 1021 GDLVDAEEYLVPOGCFPCPDPAFGAGVHHRSSSTRSGGDLTGLSESEERASPS 1080
DB 1015 GDLVDAEEYLVPOGCFPCPDPAFGAGVHHRSSSTRSGGDLTGLSESEERASPS 1074
QY 1081 LAPSEGASDVDFDGLGMAKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCS 1140
DB 1075 LAPSEGASDVDFDGLGMAKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCS 1134
QY 1141 POPEYVNOQDVRRPQPSBREGPLPAARPAGATLERAKTLSFGKGVVNDVAFGAVENP 1200
DB 1135 POPEYVNOQDVRRPQPSBREGPLPAARPAGATLERAKTLSFGKGVVNDVAFGAVENP 1194
QY 1201 EYLTPQGAAPQHPHPPAFSPAFNDLYWDDPBERGAPSTFTGTPTAENPEYLGIDVP 1260
DB 1195 EYLTPQGAAPQHPHPPAFSPAFNDLYWDDPBERGAPSTFTGTPTAENPEYLGIDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KM Therapeutic compound; major histocompatibility complex; vaccine;
KM antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KM adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KM antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 774..782
FT /note= "Antigenic epitope"
XX
XX MO200168677-A2.
XX PD 20-SEP-2001.
XX PF 16-MAR-2001; 2001MO-US40328.
XX PR 16-MAR-2000; 2000US-0527487.
XX (GENZ ) GENZYME CORP.
XX Nicoslette CA;
XX PI
XX WPI: 2001-616284/71.
XX DR N-PSDB; AAD19731.
XX
```

```
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT
XX Claim 4; Page 63-67; 69pp; English.
XX
PS The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6632; DB 22; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1226; Conservative 8; Mismatches 21; Indels 6; Gaps 1;
QY 1 MELALCRMGLLALLPFGAASVQCTGTDKRLPASPETHLDLRLHYOGCQVVGSL 60
DB 1 MELALCRMGLLALLPFGAASVQCTGTDKRLPASPETHLDLRLHYOGCQVVGSL 60
QY 61 ELTYLPTNASLSFLODIOEVGYVLIANQVROVPLQRLIRVGTOLFEDNYALAVDNG 120
DB 61 ELTYLPTNASLSFLODIOEVGYVLIANQVROVPLQRLIRVGTOLFEDNYALAVDNG 120
QY 121 DPLNNTPTVGTASPGGLRELQRLSLTBILKGGVLIQRNPOLCYODTILMKOIFHNQOLA 180
DB 121 DPLNNTPTVGTASPGGLRELQRLSLTBILKGGVLIQRNPOLCYODTILMKOIFHNQOLA 180
QY 181 LTLIDNRSRACHPCSPMKSGSRGCESSSDCQSLTRTVACAGCARCKGAPLPTDCHEOC 240
DB 181 LTLIDNRSRACHPCSPMKSGSRGCESSSDCQSLTRTVACAGCARCKGAPLPTDCHEOC 240
QY 241 AAGCTGPHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300
DB 241 AAGCTGPHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300
QY 301 YNYLSTVGSCTLVCPAHNOEVTAEQDQRECKSKPCARVCYGLMGIYIKANSKFIGIT 360
DB 301 YNYLSTVGSCTLVCPAHNOEVTAEQDQRECKSKPCARVCYGLMGIYIKANSKFIGIT 360
QY 361 ELBFAAGCKKIFGSLAFPLPESFQDPASNTAPLQPEOLQVFTLEITGYLIASAPDSL 420
DB 361 ELBFAAGCKKIFGSLAFPLPESFQDPASNTAPLQPEOLQVFTLEITGYLIASAPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAVSLTLQGLIGISWLGRLSRLREGSGGALILHNTHL 480
DB 421 DLSVFQNLQVIRGRILHNGAVSLTLQGLIGISWLGRLSRLREGSGGALILHNTHL 480
QY 481 PMDQLFRPNFTYSFMLRVKVSASHLECYGEGGLACHQLCARGHCMGPGFTQCVNCSOF 540
DB 481 PMDQLFRPNFTYSFMLRVKVSASHLECYGEGGLACHQLCARGHCMGPGFTQCVNCSOF 534
QY 541 LRGOECVBCRVLOGLPREYVNAHCLPCHECQPONGSVYTCFPEADQCVACHYXDP 600
DB 541 LRGOECVBCRVLOGLPREYVNAHCLPCHECQPONGSVYTCFPEADQCVACHYXDP 594
QY 601 FCVACRSGVYKPDLSYPIWKFPDEBGAQPCPINCTHSCVDLDDKCCPAEQRASPLTSI 660
DB 601 FCVACRSGVYKPDLSYPIWKFPDEBGAQPCPINCTHSCVDLDDKCCPAEQRASPLTSI 654
```

```
QY 661 VSAVVGILLVVLGVVFGILIKRRQOKIKRYTMRRLLOETELVEPLTPSGAMPNOQMRI 720
    |||||
DB 655 VSAVVGILLVVLGVVFGILIKRRQOKIKRYTMRRLLOETELVEPLTPSGAMPNOQMRI 714
    |||||
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPGENVKIKVAILKVLRENTSPKANKEILDEAYVM 780
    |||||
DB 715 LKETELRKVKVLGSGAFGTYYKGIWIPGENVKIKVAILKVLRENTSPKANKEILDEAYVM 774
    |||||
QY 781 AGVSPVYSRLIGICTSTVOLVTOLMPYGCILDHVRENRGLSGODLNNCMQIAKMS 840
    |||||
DB 775 AGVSPVYSRLIGICTSTVOLVTOLMPYGCILDHVRENRGLSGODLNNCMQIAKMS 834
    |||||
QY 841 YLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPKIMWALESI 900
    |||||
DB 835 YLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPKIMWALESI 894
    |||||
QY 901 LRRFTHOSDVWSYGVYTWELMTFGAKPYDGI PAEIPDLLEKGRLLPOPPICITIDVTMI 960
    |||||
DB 895 LRRFTHOSDVWSYGVYTWELMTFGAKPYDGI PAEIPDLLEKGRLLPOPPICITIDVTMI 954
    |||||
QY 961 MVKCMIDSECRPRELVSFSSRMARDPQRFVVIQNEDLGASPPLDSTFYRSLLEDDM 1020
    |||||
DB 955 MVKCMIDSECRPRELVSFSSRMARDPQRFVVIQNEDLGASPPLDSTFYRSLLEDDM 1014
    |||||
QY 1021 GDLVDAEYLVPOGFCPCPDAPGAGVNHHRSSSTRSGGDLTLGLEPSEBEAPRSP 1080
    |||||
DB 1015 GDLVDAEYLVPOGFCPCPDAPGAGVNHHRSSSTRSGGDLTLGLEPSEBEAPRSP 1074
    |||||
QY 1081 LAPBEGAGSDVFDGDLGMAKGLQSLPTHPSPLOKRSDDPTVPLBETTGYYAPLTCS 1140
    |||||
DB 1075 LAPBEGAGSDVFDGDLGMAKGLQSLPTHPSPLOKRSDDPTVPLBETTGYYAPLTCS 1134
    |||||
QY 1141 POPYVNOQDVRPOPSPREGPLPAAPAGATLERAKTLSPGKGVNVDFAFGAVENP 1200
    |||||
DB 1135 POPYVNOQDVRPOPSPREGPLPAAPAGATLERAKTLSPGKGVNVDFAFGAVENP 1194
    |||||
QY 1201 EYLTPQGGAAPQHPHPAFSPAFDNLVYWDODPEPARGAPSTFKGTPTAENPEYLGIDVP 1260
    |||||
DB 1195 EYLTPQGGAAPQHPHPAFSPAFDNLVYWDODPEPARGAPSTFKGTPTAENPEYLGIDVP 1254
    |||||
QY 1261 V 1261
    |||||
DB 1255 V 1255
    |||||

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KM Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX
OS Homo sapiens.
XX Synthetic.
XX
PN WO200100244-A2.
PD
XX 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R.
```

```
XX WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
    |||||
Query Match 96.9%; Score 6632; DB 22; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1226; Conservative 8; Mismatches 21; Indels 6; Gaps 1;

QY 1 MELALCRWGLLALLPPGAASTQVCTGDMKLRPASPTHLDMLRHLVOCQVVGNTL 60
    |||||
DB 1 MELALCRWGLLALLPPGAASTQVCTGDMKLRPASPTHLDMLRHLVOCQVVGNTL 60
    |||||
QY 61 ELTYLPTNASISFLQDIOEVGVYVLAHNOYRQVPLQRLRIVRGTLFEDNYALAVLNG 120
    |||||
DB 61 ELTYLPTNASISFLQDIOEVGVYVLAHNOYRQVPLQRLRIVRGTLFEDNYALAVLNG 120
    |||||
QY 121 DPLNNTPTVGTASPGGLRELOLRSLTELKGVLIQRNPOLCYQDTILMKDIFHNQOLA 180
    |||||
DB 121 DPLNNTPTVGTASPGGLRELOLRSLTELKGVLIQRNPOLCYQDTILMKDIFHNQOLA 180
    |||||
QY 181 LTLIDITNRSRACHPCSPMKCSRCMGSESEDCQSLTRTVCAAGCARCKGPIPTDCHEOC 240
    |||||
DB 181 LTLIDITNRSRACHPCSPMKCSRCMGSESEDCQSLTRTVCAAGCARCKGPIPTDCHEOC 240
    |||||
QY 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCVTACP 300
    |||||
DB 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCVTACP 300
    |||||
QY 301 YNYLSTDVGSCTLVCPHLNDEVTAEQDQRCCKSKPARVCYGLGMQYIANKSFIGIT 360
    |||||
DB 301 YNYLSTDVGSCTLVCPHLNDEVTAEQDQRCCKSKPARVCYGLGMQYIANKSFIGIT 360
    |||||
QY 361 ELBFAAGCKIIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEETGYLYISAMPDSL 420
    |||||
DB 361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEETGYLYISAMPDSL 420
    |||||
QY 421 DLSVFQNLQVIRGRILHNGAVSLTLQIGISWLGILRSLEIGSGLALIHNTHLCFVHTV 480
    |||||
DB 421 DLSVFQNLQVIRGRILHNGAVSLTLQIGISWLGILRSLEIGSGLALIHNTHLCFVHTV 480
    |||||
QY 481 PMDQLFFNNFTVFMILRPVVSASHLEBCEGBGLAQOLCARHGCMWPGGTQVCNQGOF 540
    |||||
DB 481 PMDQLFFNNFTVFMILRPVVSASHLEBCEGBGLAQOLCARHGCMWPGGTQVCNQGOF 534
    |||||
QY 541 LRQGECEBECVNLQGLPREVYNARHCLPCHECQPONGSVTCFGEADQCAVACHYXDP 600
    |||||
DB 541 LRQGECEBECVNLQGLPREVYNARHCLPCHECQPONGSVTCFGEADQCAVACHYXDP 594
    |||||
QY 601 FCVARCPGKVPDLSYPIWKFPEDEGACQPCPINTCHSCVLDLDDKCPAQORASPLTSI 660
    |||||
DB 595 FCVARCPGKVPDLSYPIWKFPEDEGACQPCPINTCHSCVLDLDDKCPAQORASPLTSI 654
    |||||
QY 661 VSAVVGILLVVLGVVFGILIKRRQOKIKRYTMRRLLOETELVEPLTPSGAMPNOQMRI 720
    |||||
DB 655 VSAVVGILLVVLGVVFGILIKRRQOKIKRYTMRRLLOETELVEPLTPSGAMPNOQMRI 714
    |||||
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPGENVKIKVAILKVLRENTSPKANKEILDEAYVM 780
    |||||
```

```

Db      715 LKETELKVKVILSGAGTCTVTKGIWIPDGENVKIPVAIKVLRENTSPPANKEILDEAVYM 774
Qy      781 AGVGSPPVYSRLIGICTSTVOLVTOLMPEYGLLDHVRNREGISQDILNMCQIAKMS 840
Db      775 AGVGSPPVYSRLIGICTSTVOLVTOLMPEYGLLDHVRNREGISQDILNMCQIAKMS 834
Qy      841 YLEEVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETBYHADGSKVPIKMMALESI 900
Db      835 YLEEVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETBYHADGSKVPIKMMALESI 894
Qy      901 LRRRFTQSDVMSYGVTVWELMTFGAKPYDGI PAEITDLEKGRLLGQPPICITDVMYI 960
Db      895 LRRRFTQSDVMSYGVTVWELMTFGAKPYDGI PAEITDLEKGRLLGQPPICITDVMYI 954
Qy      961 MVKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQNDLGPASPLDSTFYRSLLEDDDM 1020
Db      955 MVKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQNDLGPASPLDSTFYRSLLEDDDM 1014
Qy      1021 GDLYDAEYLVPOCGFCEPDPAFGAGVHHRHSSSTRSGGDLTTLGLEPSEEFAPRSP 1080
Db      1015 GDLYDAEYLVPOCGFCEPDPAFGAGVHHRHSSSTRSGGDLTTLGLEPSEEFAPRSP 1074
Qy      1081 LAPSEGSQDVFDGDLGMAKGLQSLPTHDPSPLORESQDPTVPLPSETDGYVAPLTCG 1140
Db      1075 LAPSEGSQDVFDGDLGMAKGLQSLPTHDPSPLORESQDPTVPLPSETDGYVAPLTCG 1134
Qy      1141 POPEYVNPQDVPRPQPSREBEPPLPARPAGATLERAKTSLPGKNGVNDVAFGAGVNP 1200
Db      1135 POPEYVNPQDVPRPQPSREBEPPLPARPAGATLERAKTSLPGKNGVNDVAFGAGVNP 1194
Qy      1201 EYLTPQGAAPQPPPPAFSPAFDNLVYWDODPPRGAPESTFKGTPFAENPEYIGLDVP 1260
Db      1195 EYLTPQGAAPQPPPPAFSPAFDNLVYWDODPPRGAPESTFKGTPFAENPEYIGLDVP 1254
Qy      1261 V 1261
Db      1255 V 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein. 1255 AA.
AC AAU74545;
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macropneagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
PD 03-JAN-2002.
PF 16-MAR-2001; 2001US-0811123.
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIKOWSKI M.
XX

```

```

PI Erickson S, Schwall R, Slikowski M,
XX
XX WPI, 2002-163686/21.
DR N-PSDB; ABK14058.
XX
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal
XX
PS Example 3; Fig 7; 93pp; English.
XX
XX The invention relates to treating a tumour in a mammal, where the tumour
XX is characterised by the overexpression of an epidermal growth factor
XX receptor (ErbB) and does not respond or responds poorly, to treatment
XX with an anti-ErbB antibody, comprising administering to the mammal an
XX anti-ErbB antibody-maytansinoid conjugate. The method is useful for
XX treating cancer or tumours of the breast, ovary, stomach, endometrium,
XX salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
XX prostate and bladder, preferably breast cancer. The breast cancer is a
XX metastatic breast cancer or an aggressive form of metastatic breast
XX cancer which overexpresses ErbB2. The method is also useful for treating
XX neuronal, glial, astrocytal, hypothalamic, glandular, macropneagal,
XX epithelial, stromal, blastocoealic, inflammatory, angiogenic and
XX immunological disorders. This sequence represents the human HER2 (ErbB2)
XX polypeptide of the invention.
XX
SQ Sequence 1255 AA;
XX
XX Query Match 96.9%; Score 6632; DB 23; Length 1255;
XX Best Local Similarity 97.2%; Pred. No. 0;
XX Matches 1226; Conservative 8; Mismatches 21; Indels 6; Gaps 1;
XX
Qy 1 MEIALALCKWGLLALLPFGAASVQCTGTDMLRLPASBETHLMDLRLHYOCQVQVQNL 60
Db 1 MEIALALCKWGLLALLPFGAASVQCTGTDMLRLPASBETHLMDLRLHYOCQVQVQNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQYVLIANNOVROVPLQRLIRVGTQLFEDNYALVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQYVLIANNOVROVPLQRLIRVGTQLFEDNYALVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYDTIIMKDI FHKNOQLA 180
Db 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYDTIIMKDI FHKNOQLA 180
Qy 181 LTLIDINRSRACHPESPWCXGSRCKGSESEDCQSLTRVCAAGCARCGPLPTDCHEQC 240
Db 181 LTLIDINRSRACHPESPWCXGSRCKGSESEDCQSLTRVCAAGCARCGPLPTDCHEQC 240
Qy 241 AAGCTGPRGSDCLAHFNHSGICEIACPALVTVNTDFEESMPNREGRTFFGASCVTACP 300
Db 241 AAGCTGPRGSDCLAHFNHSGICEIACPALVTVNTDFEESMPNREGRTFFGASCVTACP 300
Qy 301 YNVLSTVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFIGIT 360
Db 301 YNVLSTVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFIGIT 360
Qy 361 ELERAGCKKTFGSLAFIPESFDGPASNTAPLOPQLOVFETLEITGVLIYISAMPDISLP 420
Db 361 ELERAGCKKTFGSLAFIPESFDGPASNTAPLOPQLOVFETLEITGVLIYISAMPDISLP 420
Qy 421 DLSTVFONLQVIRGRIILHNAYSLTLOGIGISMLGRLSRLREGSGALALHHNTHLCFVHTV 480
Db 421 DLSTVFONLQVIRGRIILHNAYSLTLOGIGISMLGRLSRLREGSGALALHHNTHLCFVHTV 480
Qy 481 PMDQLFRPNFTVSFWLVRPVYASHSLEECVGBGLACHQLCARGHCWPGPFCVNCSCOF 540
Db 481 PMDQLFRPNFTVSFWLVRPVYASHSLEECVGBGLACHQLCARGHCWPGPFCVNCSCOF 540
Qy 541 LRGECEVEECRVLOGLFREYVNAHCLPCHBECQPOKSVTCFGBEADQCAACAHYKPP 600
Db 541 LRGECEVEECRVLOGLFREYVNAHCLPCHBECQPOKSVTCFGBEADQCAACAHYKPP 600
Qy 535 LRGECEVEECRVLOGLFREYVNAHCLPCHBECQPOKSVTCFGBEADQCAACAHYKPP 594
Db 535 LRGECEVEECRVLOGLFREYVNAHCLPCHBECQPOKSVTCFGBEADQCAACAHYKPP 594
Qy 601 FCVARCPSGVYKPDLSYMPIMKFPDEBAGACQCPINCHSGCVTDLDKCGPARQASPLTISI 660
Db 601 FCVARCPSGVYKPDLSYMPIMKFPDEBAGACQCPINCHSGCVTDLDKCGPARQASPLTISI 660

```

```

Db      |||||
595  FCVACPSGVKPDLSMPIMKFPDEGACOPCPINCTHSCVLDLDDKGCPCAEQASPLTSI 654
Qy      |||||
661  VSAVVGILLVVLGVVFGILIKRROOKIKKYMRLLOETELVEPIITPGAMPNOQM1 720
Db      |||||
655  VSAVVGILLVVLGVVFGILIKRROOKIKKYMRLLOETELVEPIITPGAMPNOQM1 714
Qy      |||||
721  LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLENTSPKANKETLDEAYM 780
Db      |||||
715  LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLENTSPKANKETLDEAYM 774
Qy      |||||
761  AGVSPVYSRLIGICTSTVOLVTLQMPYCGCLLDHVENRGRLGSDLLNMCMQIAKMS 840
Db      |||||
775  AGVSPVYSRLIGICTSTVOLVTLQMPYCGCLLDHVENRGRLGSDLLNMCMQIAKMS 834
Qy      |||||
841  YLEDVRLVHRDLAARVNLKSPNHVKITDPSGLARLLDIDETEHAAQGVKPKMMALESI 900
Db      |||||
835  YLEDVRLVHRDLAARVNLKSPNHVKITDPSGLARLLDIDETEHAAQGVKPKMMALESI 894
Qy      |||||
901  LRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPOPICTTIDVYI 960
Db      |||||
895  LRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPOPICTTIDVYI 954
Qy      |||||
961  MWKCMWIDSECRPRRELVSERFSKMARDPQRFVIONEDLGPASPLDSTFYSLLEDDM 1020
Db      |||||
955  MWKCMWIDSECRPRRELVSERFSKMARDPQRFVIONEDLGPASPLDSTFYSLLEDDM 1014
Qy      |||||
1021  GDLVAAEEYLPQOGFPCDPAPAGAGMWHHRSSSTSGGDLTLGLPESEEPSP 1080
Db      |||||
1015  GDLVAAEEYLPQOGFPCDPAPAGAGMWHHRSSSTSGGDLTLGLPESEEPSP 1074
Qy      |||||
1081  LAPSEGASDVEFDGLGMGAAGLQSLPTHDSPLQRYSEDPVPLPSETDGVAPLTS 1140
Db      |||||
1075  LAPSEGASDVEFDGLGMGAAGLQSLPTHDSPLQRYSEDPVPLPSETDGVAPLTS 1134
Qy      |||||
1141  POPEYVNOPDVAPPOPPSPREGPLPAARPGATLERAKTISPGKNGVYKOVFAFGAVENP 1200
Db      |||||
1135  POPEYVNOPDVAPPOPPSPREGPLPAARPGATLERAKTISPGKNGVYKOVFAFGAVENP 1194
Qy      |||||
1201  EYLTPOGGAPOPPPPAPAFNDLYYWDOPPERGAPSTPFKGTPTAENEYVGLDVP 1260
Db      |||||
1195  EYLTPOGGAPOPPPPAPAFNDLYYWDOPPERGAPSTPFKGTPTAENEYVGLDVP 1254
Qy      |||||
1261  V 1261
Db      |||||
1255  V 1255

```

RESULT 5  
AAW01111  
ID AAW01111 standard; Protein; 1255 AA.  
AC AAW01111;

```

DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu: c-erbB1; p185; oncogene; tyrosine protein kinase;
KM breast cancer; ovary cancer; colon cancer; lung cancer;
KK prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 676..1255
FT Domain
FT /label= Intracellular_domain
FT /note= "claimed domain, useful for immunisation"
XX
PN MO9630514-A1.
XX
PD 03-OCT-1996.
XX

```

```

PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNITV ) UNITV WASHINGTON.
XX
PI Cheever MA, Distle ML;
XX
DR WPI: 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2/neu polypeptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2, Page 56-61; 71pp; English.
XX
CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
XX
Query Match 96.8%; Score 6626; DB 17; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
Qy 1 MELALACRWGLLALLPGAASTOYCTGDMKRLPASPETHLDMLRHLYOGCOVVGNI 60
Db 1 MELAALCRWGLLALLPGAASTOYCTGDMKRLPASPETHLDMLRHLYOGCOVVGNI 60
Qy 61 ELTYVPTNASSLFLODIOVGQYVLIANQVAVQVLPQRIYRGQLPFDNALAVLDNG 120
Db 61 ELTYVPTNASSLFLODIOVGQYVLIANQVAVQVLPQRIYRGQLPFDNALAVLDNG 120
Qy 121 DELNNTTPYVGASPGGLRELQRLSLTEILKGVLIQRNPQLCYODTILMKDIFHKNOQL 180
Db 121 DELNNTTPYVGASPGGLRELQRLSLTEILKGVLIQRNPQLCYODTILMKDIFHKNOQL 180
Qy 181 LTLIDTNSRACHPCSPMKCKSGRCESESDQSLTRTYCAGGACARCKPLPTDCHEOC 240
Db 181 LTLIDTNSRACHPCSPMKCKSGRCESESDQSLTRTYCAGGACARCKPLPTDCHEOC 240
Qy 241 AAGCTGPKHSDCLALCHFHSGICELHCPALVTYNTDFESMPNBEGRYTFASCTYACP 300
Db 241 AAGCTGPKHSDCLALCHFHSGICELHCPALVTYNTDFESMPNBEGRYTFASCTYACP 300
Qy 301 YNYLSTDVSGCTLVCPILHNOEVTADGTQRCESKSPCARVCYGLQMOYIKANSKFIGIT 360
Db 301 YNYLSTDVSGCTLVCPILHNOEVTADGTQRCESKSPCARVCYGLQMOYIKANSKFIGIT 360
Qy 361 ELIEFAGCKKIFGSLAFLPESPDGPASNTAPLOPOLQVFEETLEITGVLYISAMPDSIP 420
Db 361 IOEFAGCKKIFGSLAFLPESPDGPASNTAPLOPOLQVFEETLEITGVLYISAMPDSIP 420
Qy 421 DLSVFQNTQVIRGRILHNGAYSLTQIGISMLGRLSRLREIGSGALILHNNHLCFVHV 480
Db 421 DLSVFQNTQVIRGRILHNGAYSLTQIGISMLGRLSRLREIGSGALILHNNHLCFVHV 480
Qy 481 FMDQLFRNNFTVSMRLVFPKVASHLSECVGEGLAGLQLCARGHGKMGFPQCVNCSQF 540
Db 481 FMDQLFRNNFTVSMRLVFPKVASHLSECVGEGLAGLQLCARGHGKMGFPQCVNCSQF 540
Qy 541 LRGECEVCEKVLQGLPREYVNAARCLPCHPECOPQNSVTCFGEADQCAAHYKDP 600
Db 541 LRGECEVCEKVLQGLPREYVNAARCLPCHPECOPQNSVTCFGEADQCAAHYKDP 600

```

```

QY 601 FCVACPSGVKPDLSYMPIMKFPDEBGAOPCPINCTHS CVLDLDDKCPAEORASPLTSTI
DB 595 FCVACPSGVKPDLSYMPIMKFPDEBGAOPCPINCTHS CVLDLDDKCPAEORASPLTSTI
QY 661 VSAVVGILLVVLGVVFGILIKRROOKIKRYTMRILQETELVEPLTPSGAMPNQOMI
DB 655 ISAVVGILLVVLGVVFGILIKRROOKIKRYTMRILQETELVEPLTPSGAMPNQOMI
QY 721 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKELIDEA YVM
DB 715 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKELIDEA YVM
QY 761 AGVSPVYSRLIGICLTSTVOLVTOUMPGCLLDHRENRGRIGSGDILNMCQIAKGM
DB 775 AGVSPVYSRLIGICLTSTVOLVTOUMPGCLLDHRENRGRIGSGDILNMCQIAKGM
QY 841 YLEDVRLVHRDLAARVNLVKS PNHVKITD FGLARLIDIDETEHADGKVP I KMMALESI
DB 835 YLEDVRLVHRDLAARVNLVKS PNHVKITD FGLARLIDIDETEHADGKVP I KMMALESI
QY 901 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVMI
DB 895 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVMI
QY 961 MWKCMIDSECPRRRELVSERSMARDPQRFVVIQNEDLGPASPLDSTFFYSLLDDDM
DB 955 MWKCMIDSECPRRRELVSERSMARDPQRFVVIQNEDLGPASPLDSTFFYSLLDDDM
QY 1021 GDLVDAEYLVPQGFPCPDPAAGAGMWHHRSSSTSSGGGDLTLGLPSEBEAPRSP
DB 1015 GDLVDAEYLVPQGFPCPDPAAGAGMWHHRSSSTSSGGGDLTLGLPSEBEAPRSP
QY 1081 LAPSEGASDVDFDGLGMAKAGLQSLPTHDPSPLOQYSEDPVPLPSETIDGVAPLTCS
DB 1075 LAPSEGASDVDFDGLGMAKAGLQSLPTHDPSPLOQYSEDPVPLPSETIDGVAPLTCS
QY 1141 POPEVYNQDVNPOPPSPREPLPAARPAKATLEAKTISPGKNGVKNVFAFGAVENP
DB 1135 POPEVYNQDVNPOPPSPREPLPAARPAKATLEAKTISPGKNGVKNVFAFGAVENP
QY 1201 EYLTPQGAAPQHPHPAPSPAFNDLYYNDQDPERGAPSTFGKPTAENEYGLDVP
DB 1195 EYLTPQGAAPQHPHPAPSPAFNDLYYNDQDPERGAPSTFGKPTAENEYGLDVP
QY 1261 V 1261
DB 1255 V 1255

```

RESULT 6  
AAW92406  
ID AAW92406 standard; Protein; 1255 AA.

```

AC AAW92406;
DT 21-APR-1999 (first entry)
DE Human HER-2/neu oncogene protein.
EE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
EE malignancy; treatment; tumour.
OS Homo sapiens.
XX Key Location/Qualifiers
XX Region 676..1255
XX /note= "region which elicits immune response"
XX US5869445-A.
XX 09-FEB-1999.
XX 01-APR-1996; 96US-0625101.

```

```

XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX (UNIT ) UNIV WASHINGTON.
XX Cheever MA, Dist ML;
XX MPI: 1999-152835/13.
XX N-PSDB; AAX01912.
XX Use of HER-2/neu polypeptides - for eliciting an immune response to
XX an HER-2/neu associated malignancy, particularly for treating or
XX preventing tumours
XX Claim 3; Column 31-38; 26pp; English.
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
XX of this protein is used in a method for eliciting or enhancing an immune
XX response to HER-2/neu protein. The polypeptide can stimulate T cells and
XX B cells to produce an immune response to the HER-2/neu protein. The
XX method can be used for immunisation against a malignancy in which the
XX HER-2/neu oncogene is associated and in the treatment of an existing
XX tumour, or to prevent tumour occurrence or recurrence.
SQ Sequence 1255 AA;
Query Match 96.8%; Score 6626; DB 20; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
QY 1 METALACRMGILLALIPGASTOYCTGDMRLRPAPEYTHLDMRLHYOGGVQGNL 60
DB 1 MELALACRMGILLALIPGASTOYCTGDMRLRPAPEYTHLDMRLHYOGGVQGNL 60
QY 61 ELTYLPTNASLSFLDIOEVGVVLI AHNVQVPLQRIIRYRGTLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVGVVLI AHNVQVPLQRIIRYRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYODTILMKDIFHXNQOLA 180
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYODTILMKDIFHXNQOLA 180
QY 181 LTLIDNRSRACHPCSPCKSGRCSGSESDQSLTRTYCAGGACRCKPLPTDCCHEOC 240
DB 181 LTLIDNRSRACHPCSPCKSGRCSGSESDQSLTRTYCAGGACRCKPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTPESMPNPEGRTYFCASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTPESMPNPEGRTYFCASCVTACP 300
QY 301 YNYISTDVGSCTLVCP LHNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFTIGIT 360
DB 301 YNYISTDVGSCTLVCP LHNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFTIGIT 360
QY 361 ELEFAGCKKIRGSLAFPLESPDGPASNTAPLPQPOLQVFFELTEITGLYISAMPDLSLP 420
DB 361 ELEFAGCKKIRGSLAFPLESPDGPASNTAPLPQPOLQVFFELTEITGLYISAMPDLSLP 420
QY 421 DLSVFQNTQVIRGRILHNGAVSLTLOGIGISWLGRLSRLRELSGLALHHNTHLCFVHTV 480
DB 421 DLSVFQNTQVIRGRILHNGAVSLTLOGIGISWLGRLSRLRELSGLALHHNTHLCFVHTV 480
QY 481 PWDOLFRRNNTVTSFWLVRPKVASHLEECVGEGLACHQLCARHGCWGGPTQVCNCSQF 540
DB 481 PWDOLFRRNNTVTSFWLVRPKVASHLEECVGEGLACHQLCARHGCWGGPTQVCNCSQF 540
QY 541 LRGOECVEECRVLOGLPREYNNARHCLPCHRECOPONSVCYCFGEAQQCAVCAHYKDP 600
DB 541 LRGOECVEECRVLOGLPREYNNARHCLPCHRECOPONSVCYCFGEAQQCAVCAHYKDP 600
QY 594 LRGOECVEECRVLOGLPREYNNARHCLPCHRECOPONSVCYCFGEAQQCAVCAHYKDP 594

```

```

QY 601 FCVARCPSGVKBDLSYMPIWKPEDEGACQPCPINCTHSCVDLDDKGCRAEQRASPLTGI 660
DB 595 FCVARCPSGVKBDLSYMPIWKPEDEGACQPCPINCTHSCVDLDDKGCRAEQRASPLTGI 654
QY 661 VSAVVGILLVVLGVVFGILLIKRROOKIRKTYTKRLLQETELVEPITPSGAMPNQOMRI 720
DB 655 ISAVVGILLVVLGVVFGILLIKRROOKIRKTYTKRLLQETELVEPITPSGAMPNQOMRI 714
QY 721 LKETELRKVKVLGSGAFGTVYKGIWI PDGENVKI PAIVKYLRENTSPKXNKELLDAAVYM 780
DB 715 LKETELRKVKVLGSGAFGTVYKGIWI PDGENVKI PAIVKYLRENTSPKXNKELLDAAVYM 774
QY 781 AGVGSFYVSRLLGICLTSTVQVLTQMLPYGCLLDHRENGRIGSODLLNMCQIAKGN 840
DB 775 AGVGSFYVSRLLGICLTSTVQVLTQMLPYGCLLDHRENGRIGSODLLNMCQIAKGN 834
QY 841 YLEDVRLVHRDLAARVNLVKSPPHVKITTFGLARLLDIDETEHADGKVPKIMMALESI 900
DB 835 YLEDVRLVHRDLAARVNLVKSPPHVKITTFGLARLLDIDETEHADGKVPKIMMALESI 894
QY 901 LRRRFTHSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPCITIDVYMI 960
DB 895 LRRRFTHSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPCITIDVYMI 954
QY 961 MYKCMWIDSECRPRFRELVSFESRMARDQRFVVIQNEIDGPSASPLDSTFYRSLLEDDEM 1020
DB 955 MYKCMWIDSECRPRFRELVSFESRMARDQRFVVIQNEIDGPSASPLDSTFYRSLLEDDEM 1014
QY 1021 GLVDAEELVLPQOGFCFCDPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEEA PRSP 1080
DB 1015 GLVDAEELVLPQOGFCFCDPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEEA PRSP 1074
QY 1081 LAPSEGASDVDFGDLGMAAGKLOSLPTHDPSP LRYSEDPVPLPSETDGVVAULTGS 1140
DB 1075 LAPSEGASDVDFGDLGMAAGKLOSLPTHDPSP LRYSEDPVPLPSETDGVVAULTGS 1134
QY 1141 POREYVNOQDVPRQPPSPREGPLPAARPAATLERAKTSLPGKNGVVKDVFAGAVENP 1200
DB 1135 POREYVNOQDVPRQPPSPREGPLPAARPAATLERAKTSLPGKNGVVKDVFAGAVENP 1194
QY 1201 EYLTQGGAAPOHPHPAPSPAFDNLTYNDODPEREGAPSPSTFKGPTAENPEYLGLDVP 1260
DB 1195 EYLTQGGAAPOHPHPAPSPAFDNLTYNDODPEREGAPSPSTFKGPTAENPEYLGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255

```

RESULT 7  
AAB21198  
ID AAB21198 standard; protein; 1255 AA.  
XX AAB21198;  
AC AAB21198;  
XX AAB21198;  
DT 12-JAN-2001 (first entry)  
XX 12-JAN-2001 (first entry)  
DE Human HER-2/neu protein.  
XX Human HER-2/neu protein.  
KM Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;  
XX breast cancer; prostate cancer; ovarian cancer; lung cancer;  
XX colon cancer.  
XX Homo sapiens.  
XX OS  
XX WO200044899-A1.  
XX 03-AUG-2000.  
XX 28-JAN-2000; 2000WO-US02164.  
XX PF  
XX 29-JAN-1999; 99US-0117976.  
XX

```

PA (CORI-) CORIXA CORP.
PA (SMIK ) SMITHKLINE BEECHAM.
XX Cheever MA, Gheyssen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX Claim 52; Fig 7; 128pp; English.
XX The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
SQ Sequence 1255 AA;
Query Match 96.8%; Score 6626; DB 21; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
QY 1 MELALCRGGLLALLPPGAASVCTGTDMKRLRLPASFEHLDMRLHYOGGVQGNL 60
DB 1 MELALCRGGLLALLPPGAASVCTGTDMKRLRLPASFEHLDMRLHYOGGVQGNL 60
QY 61 EYLTPTNASLFLDIOGVGVVLIANQVQVPLQRIYRGTOLEFDNVALAVLDNG 120
DB 61 EYLTPTNASLFLDIOGVGVVLIANQVQVPLQRIYRGTOLEFDNVALAVLDNG 120
QY 121 DPLANTPTVYTGASPGGLRELQRLSLTEILKGVLIQRPQLCYQDITLKKDIFHXNOLA 180
DB 121 DPLANTPTVYTGASPGGLRELQRLSLTEILKGVLIQRPQLCYQDITLKKDIFHXNOLA 180
QY 181 LTLIDTNRBACHPGSPMKSGRCWGESSEDCQSLTRYTCAGGCARCKGRLPTDCCHEQC 240
DB 181 LTLIDTNRBACHPGSPMKSGRCWGESSEDCQSLTRYTCAGGCARCKGRLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHPCALVTYNTDFESMPNPRGRTTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHPCALVTYNTDFESMPNPRGRTTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPVLHNOEYTABDGTORCEKSPKCARVCYGLGMQYIKANSKEFIGIT 360
DB 301 YNYLSTDVSGCTLVCPVLHNOEYTABDGTORCEKSPKCARVCYGLGMQYIKANSKEFIGIT 360
QY 361 EYLFAGCKKIFGSLAFLEPSEFDGDPASNTAPQPEBOLQVFETLEBITGVLISAMPDLSIP 420
DB 361 EYLFAGCKKIFGSLAFLEPSEFDGDPASNTAPQPEBOLQVFETLEBITGVLISAMPDLSIP 420
QY 421 DLSVFONLQVIRGLIHNGAYSLTQGLGISWLGRLSRLBELSGALIHNNHILCFVHV 480
DB 421 DLSVFONLQVIRGLIHNGAYSLTQGLGISWLGRLSRLBELSGALIHNNHILCFVHV 480
QY 481 PWDOLFERNFTVSEFVLVPRKVSASHLECEVGEGLACHQLCARGHCWGPGPTQCVVCSOF 540
DB 481 PWDOLFERNFTVSEFVLVPRKVSASHLECEVGEGLACHQLCARGHCWGPGPTQCVVCSOF 534
QY 541 LRGOCEVECRVLOGLPREYVNAARHCLPHRECCQPNQSVYTFGGPADOCVACAHYKDP 600
DB 541 LRGOCEVECRVLOGLPREYVNAARHCLPHRECCQPNQSVYTFGGPADOCVACAHYKDP 594
QY 601 FCVARCPSGVKBDLSYMPIWKPEDEGACQPCPINCTHSCVDLDDKGCRAEQRASPLTGI 660

```



```
Db      |||||||
595 FCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCHTSCVDLDDKCGAERASPLTISI 654
Qy      |||||||
661 USAVVGILLVVLGVVFGILLKRRROOKIRKTYMRRLQETVELVEPLTPSGAMPNOAKRI 720
Db      |||||||
655 ISAVVGILLVVLGVVFGILLKRRROOKIRKTYMRRLQETVELVEPLTPSGAMPNOAKRI 714
Qy      |||||||
721 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDEAYVM 780
Db      |||||||
715 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDEAYVM 774
Qy      |||||||
781 AGVSPYVSRLLGICLTSTVOLVTOLMPYGCILDHVRENRGLSGODLNMCMQIAKWS 840
Db      |||||||
775 AGVSPYVSRLLGICLTSTVOLVTOLMPYGCILDHVRENRGLSGODLNMCMQIAKWS 834
Qy      |||||||
841 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGGKVPKMMALESI 900
Db      |||||||
835 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGGKVPKMMALESI 894
Qy      |||||||
901 LRRPFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
Db      |||||||
895 LRRPFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
Qy      |||||||
961 MVKCMIDSECRPRRELVSERSMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDDM 1020
Db      |||||||
955 MVKCMIDSECRPRRELVSERSMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDDM 1014
Qy      |||||||
1021 GDLYDAEEYLPQOGFCPPDPAPGAGVWVHHRSSSTRSGGDDTLGLEPSEEAAPSP 1080
Db      |||||||
1015 GDLYDAEEYLPQOGFCPPDPAPGAGVWVHHRSSSTRSGGDDTLGLEPSEEAAPSP 1074
Qy      |||||||
1081 LAPSEGASDVFDGDLGMAAKGLQSLPTHDSPLOKRYSEDPVLPSETDGVAPLTCIS 1140
Db      |||||||
1075 LAPSEGASDVFDGDLGMAAKGLQSLPTHDSPLOKRYSEDPVLPSETDGVAPLTCIS 1134
Qy      |||||||
1141 POPRYVNPQVPRPQPSRREGPLPAARAGATLERAKTLSPKKNVYKQVDFAFGAVENP 1200
Db      |||||||
1135 POPRYVNPQVPRPQPSRREGPLPAARAGATLERAKTLSPKKNVYKQVDFAFGAVENP 1194
Qy      |||||||
1201 EYLTPOGGAAPQPPAPFSPAFDLYYWDODPPERGAPESTFKGPTPAENPEYLGIDVP 1260
Db      |||||||
1195 EYLTPOGGAAPQPPAPFSPAFDLYYWDODPPERGAPESTFKGPTPAENPEYLGIDVP 1254
Qy      |||||||
1261 V 1261
Db      |||||||
1255 V 1255

RESULT 8
A184780
ID      AAY84780 standard; Protein, 1255 AA.
XX
AC      AAY84780;
XX
DT      08-AUG-2000 (first entry)
XX
DE      Amino acid sequence of the SPLICE erdb-2 receptor protein.
XX
KM      SPLICE erdb-2 receptor protein; cell transformation disorder; cancer;
KM      tumor cell proliferation; tissue degeneration; arthropathy;
KM      bone resorption; inflammatory disease; degenerative disorder;
KM      wound healing.
XX
OS      Homo sapiens.
XX
PN      WO200020579-A1.
PD      13-APR-2000.
XX
PF      01-OCT-1999; 99WO-CA00912.
PR      02-OCT-1998; 98US-0165192.
XX
```

```
PA      (UNYC-) UNIV MCMMASTER.
XX
PI      Muller WJ, Siegel PM;
XX
DR      WPI. 2000-303768/26.
XX
DR      N-PSDB; AAI14812.
XX
PT      Nucleic acid encoding an erdb 2 receptor protein designated SPLICE
XX      erdb-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS      Claim 3; Fig 2; 60pp; English.
XX
CC      The present sequence represents a SPLICE erdb-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erdb-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPLICE erdb-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPLICE erdb-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPLICE erdb-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
XX
SQ      Sequence 1255 AA;
XX
Query Match 96.8%; Score 6626; DB 21; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
Qy      1 MEIALACWGLIALLPAGASTOVCCTGDKMLRPAPEHLLMLRLHYOCQOVQGNL 60
Db      1 MEIALACWGLIALLPAGASTOVCCTGDKMLRPAPEHLLMLRLHYOCQOVQGNL 60
Qy      61 ELTYLPTNASTPLQDIOEVQGVYLIANOVROVPLQRLRIYRGTQLFEDNALAVLNG 120
Db      61 ELTYLPTNASTPLQDIOEVQGVYLIANOVROVPLQRLRIYRGTQLFEDNALAVLNG 120
Qy      121 DPLNNTTPVTGASPGGLRELOLRSLTEILKSGVLIQRNPOLCYODTILMKDIFHNKOLA 180
Db      121 DPLNNTTPVTGASPGGLRELOLRSLTEILKSGVLIQRNPOLCYODTILMKDIFHNKOLA 180
Qy      181 LTLIDTNRSRACHPCSPWCKSGRCGSESEDCQSLTRVCAAGCARGPLPTDCCHQC 240
Db      181 LTLIDTNRSRACHPCSPWCKSGRCGSESEDCQSLTRVCAAGCARGPLPTDCCHQC 240
Qy      241 AAGCTGPGHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
Db      241 AAGCTGPGHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPHLNQEVTABDGTQRCCKSKPCARVYCYGLMOYIRANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPHLNQEVTABDGTQRCCKSKPCARVYCYGLMOYIRANSKFIGIT 360
Qy      361 ELERAGCKKIRGSLAFIPESFDGPASTATLQEPOLQVFTLEITGYLISAMPDLP 420
Db      361 IQEFAAGCKKIRGSLAFIPESFDGPASTATLQEPOLQVFTLEITGYLISAMPDLP 420
Qy      421 DLSVFQNLQVIRGILHNHGAVALTLQGLIGLWLGRLSRLREGSGGALALHNHTHLCFYHTV 480
Db      421 DLSVFQNLQVIRGILHNHGAVALTLQGLIGLWLGRLSRLREGSGGALALHNHTHLCFYHTV 480
Qy      481 PMDQLFRPNFTVSFWLVRPVVSASHDEECVGBGLACHQLCARGHGMGPGPTOCVNCQSF 540
Db      481 PMDQLFRPNFTVSFWLVRPVVSASHDEECVGBGLACHQLCARGHGMGPGPTOCVNCQSF 534
Qy      541 LRGOECVEECRVLOGLPREYVNAHCLPCHECCQPNQSGSYTCRFPBADQCYACAHYKDP 600
Db      541 LRGOECVEECRVLOGLPREYVNAHCLPCHECCQPNQSGSYTCRFPBADQCYACAHYKDP 594
Qy      601 FCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCHTSCVDLDDKCGAERASPLTISI 660
```



```
Db      |||||
595 FCVACPSGVPKDLSTMPIMKPPDEGACQPCPINCTHSCVDLDKGCBAEORASPLTSI 654
Qy      |||||
661 USAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPTLPGAMPNOQMKI 720
Db      |||||
655 ISAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPTLPGAMPNOQMKI 714
Qy      |||||
721 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIVLRENTSPKANKETLDEAYVM 780
Db      |||||
715 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIVLRENTSPKANKETLDEAYVM 774
Qy      |||||
761 AGVGSPPYVSRLLIGICTLSTVOLVTOLMPYGCCLLDHVRNRRGLSODLLNMCMQIAKGS 840
Db      |||||
775 AGVGSPPYVSRLLIGICTLSTVOLVTOLMPYGCCLLDHVRNRRGLSODLLNMCMQIAKGS 834
Qy      |||||
841 YLEDRLVHRDLAARNVLVKSNNHYKITDFGLARLLDIDETEHADGKGVPIKMMALESI 900
Db      |||||
835 YLEDRLVHRDLAARNVLVKSNNHYKITDFGLARLLDIDETEHADGKGVPIKMMALESI 894
Qy      |||||
901 LRRRFTHOSDVMVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMI 960
Db      |||||
895 LRRRFTHOSDVMVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMI 954
Qy      |||||
961 MYKCMIMIDSECRPRRELVSERSMARDPQRFVITQNEIDLGPASPLDSTFYRLSLEDMD 1020
Db      |||||
955 MYKCMIMIDSECRPRRELVSERSMARDPQRFVITQNEIDLGPASPLDSTFYRLSLEDMD 1014
Qy      |||||
1021 GDLVDAEFLVPOQGFPCDPAPAGAGMHHRRSSSTSGGDLTLGLSPSEERPRSP 1080
Db      |||||
1015 GDLVDAEFLVPOQGFPCDPAPAGAGMHHRRSSSTSGGDLTLGLSPSEERPRSP 1074
Qy      |||||
1081 LAPSEGAGSDVDFGDLGMAAKGLOSLPTHDPSPLOQYSEDPVPLPSTEDGVAPLTS 1140
Db      |||||
1075 LAPSEGAGSDVDFGDLGMAAKGLOSLPTHDPSPLOQYSEDPVPLPSTEDGVAPLTS 1134
Qy      |||||
1141 POPEYVNOPDVAPPOPPSPREGPLPAARPGATLERAKTILSPGKNGVKDVFAFGAVENP 1200
Db      |||||
1135 POPEYVNOPDVAPPOPPSPREGPLPAARPGATLERAKTILSPGKNGVKDVFAFGAVENP 1194
Qy      |||||
1201 EYLTQGGAPQPPHPPAPAFDNLVYWDODPPERGAPSPSTFKGPTAENBEYLGIDVP 1260
Db      |||||
1195 EYLTQGGAPQPPHPPAPAFDNLVYWDODPPERGAPSPSTFKGPTAENBEYLGIDVP 1254
Qy      |||||
1261 V 1261
Db      |||||
1255 V 1255

RESULT 9
AAB85458
ID      AAB85458 standard; Protein; 1255 AA.
AC      AAB85458;
DT      25-SEP-2001 (first entry)
XX      Human HER-2/neu protein.
DE      Human HER-2/neu protein.
XX      Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW      oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
OS      Homo sapiens.
XX      MO200153463-A2.
XX      FN      26-JUL-2001.
XX      PD      19-JAN-2001; 2001WO-US01850.
XX      PR      21-JAN-2000; 2000US-0177545.
XX      PA      (CORI-) CORIXA CORP.
```

```
PI      Cheever MA, Hand-Zimmermann S;
XX      MPI: 2001-476112/51.
DR      N-PSDB; AAH23392.
XX      PT      New antigen-presenting cells, useful as vaccines for eliciting or
XX      PT      enhancing an immune response to HER-2/neu protein, particularly useful
XX      PT      for treating or preventing cancer, e.g. breast cancer -
XX      PS      Claim 2, Page 41-46; 49pp; English.
XX      CC      The invention provides an isolated antigen-presenting cell, which
XX      CC      expresses at least an immunogenic portion of a polypeptide that produces
XX      CC      an immune response to HER-2/neu protein. The antigen-presenting cells are
XX      CC      useful as vaccines for eliciting or enhancing an immune response to
XX      CC      HER-2/neu protein, particularly in treating or preventing malignancies in
XX      CC      which the HER-2/neu oncogene is associated. Specifically, these are
XX      CC      useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX      CC      colon, lung or prostate cancers. The present sequence represents
XX      CC      the human HER-2/neu protein (also known as p185 or c-erbB2).
```

SO Sequence 1255 AA;

Query Match 96.8%; Score 6626; DB 22; Length 1255;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;

```
Qy      1 MELALCRWGLLALLPPAASTOYCTGDMKRLPLASPTLDMRLHYOGCOVVOGSL 60
Db      1 MELALCRWGLLALLPPAASTOYCTGDMKRLPLASPTLDMRLHYOGCOVVOGSL 60
Qy      61 ELTYLPTNASLFLDIOEVQGVYLIAHQVQVPLQRIYRGTQLFEDNALAVLDNG 120
Db      61 ELTYLPTNASLFLDIOEVQGVYLIAHQVQVPLQRIYRGTQLFEDNALAVLDNG 120
Qy      121 DPLNNTPTVYTGASPGGLRELQRLSTELIKGVLIOBNPOLCYODTILMKDIFHKNOQLA 180
Db      121 DPLNNTPTVYTGASPGGLRELQRLSTELIKGVLIOBNPOLCYODTILMKDIFHKNOQLA 180
Qy      181 LTLITNNSRAHPCSPGKSGRCSGSESDQSLTRYCAGGACARCKPRLPTDCHEOC 240
Db      181 LTLITNNSRAHPCSPGKSGRCSGSESDQSLTRYCAGGACARCKPRLPTDCHEOC 240
Qy      241 AAGCTGPKHSDCLACHFNHSGICELHCAALVTYNTDFESMNPNEGRTTPGASCTYAP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICELHCAALVTYNTDFESMNPNEGRTTPGASCTYAP 300
Qy      301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Qy      361 ELIEFAGCKKIFGSLAFLESPFGDPASNTAPLQPELOVFEFLBEITGYLYISAMPDSLP 420
Db      361 IOEFAGCKKIFGSLAFLESPFGDPASNTAPLQPELOVFEFLBEITGYLYISAMPDSLP 420
Qy      421 DLSVFQNLQVIRGRLIHNGAVSLTLQGLISMLGLRSILRELGSGALIIHNHNLGVPHV 480
Db      421 DLSVFQNLQVIRGRLIHNGAVSLTLQGLISMLGLRSILRELGSGALIIHNHNLGVPHV 480
Qy      481 PWDQLFRNPFVSFWLVPKVASHLBECVGEGLAQHLCARGHCWGGPTQCVNCSQF 540
Db      481 PWDQLFRNPFVSFWLVPKVASHLBECVGEGLAQHLCARGHCWGGPTQCVNCSQF 534
Qy      541 LRQDECEBECRYLQGLPREYVNAHRLPCHPECOPONGSVTCFGEADQCVACAHYKDP 600
Db      541 LRQDECEBECRYLQGLPREYVNAHRLPCHPECOPONGSVTCFGEADQCVACAHYKDP 594
Qy      601 FCVACPSGVPKDLSTMPIMKPPDEGACQPCPINCTHSCVDLDKGCBAEORASPLTSI 660
Db      595 FCVACPSGVPKDLSTMPIMKPPDEGACQPCPINCTHSCVDLDKGCBAEORASPLTSI 654
Qy      661 USAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPTLPGAMPNOQMKI 720
Db      655 FCVACPSGVPKDLSTMPIMKPPDEGACQPCPINCTHSCVDLDKGCBAEORASPLTSI 654
```

```

Db      655 ISAVVGLLVVIGVFGILLKRRQOKIRKYMRLQETELVPLTPSGAMPNQAOMRI 714
Qy      721 LKETELRKVKVLSGAGFVYKGIIMPDGENVKIPVAIKVLRENTSPKANEIIDEAYVM 780
      715 LKETELRKVKVLSGAGFVYKGIIMPDGENVKIPVAIKVLRENTSPKANEIIDEAYVM 774
Qy      781 AGVGSPPVYSRLGLICTSTVQVLTQMPYGCGLDHVRENRGLSQDILNNCMQIAKMS 840
      775 AGVGSPPVYSRLGLICTSTVQVLTQMPYGCGLDHVRENRGLSQDILNNCMQIAKMS 834
Qy      841 YLEEVRLVHRDLAARNVLYKSPNHNKITTDFGLARLLIDETEHYADGKVPYIKWMALESI 900
      835 YLEEVRLVHRDLAARNVLYKSPNHNKITTDFGLARLLIDETEHYADGKVPYIKWMALESI 894
Qy      901 LRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPICTIDVMI 960
      895 LRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPICTIDVMI 954
Qy      961 MVKCMIDSECRPRFRELVSFSSMARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDDM 1020
      955 MVKCMIDSECRPRFRELVSFSSMARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDDM 1014
Qy      1021 GDLVDAEYLVPOGFFCPRDPAPAGGVNHRHSSSTRSGGDLTLGLPSEBEAPRSP 1080
      1015 GDLVDAEYLVPOGFFCPRDPAPAGGVNHRHSSSTRSGGDLTLGLPSEBEAPRSP 1074
Qy      1081 LAPSEGAQDVFDGDLGKAAGKGLQSLPTHDPSPLOXYSEDPTVPLPSETGYVAPLTCSS 1140
      1075 LAPSEGAQDVFDGDLGKAAGKGLQSLPTHDPSPLOXYSEDPTVPLPSETGYVAPLTCSS 1134
Qy      1141 POPEYVNPQVPRPQPSPREGPPPLAARPAATLERAKTLSPGKGVNVDAVFAFGAVENP 1200
      1135 POPEYVNPQVPRPQPSPREGPPPLAARPAATLERAKTLSPGKGVNVDAVFAFGAVENP 1194
Qy      1201 EYLTPOGGAAPQPPPPAFSPAFNLYYWDPPREKGPSTFGTPTAENPEYLGIDVP 1260
      1195 EYLTPOGGAAPQPPPPAFSPAFNLYYWDPPREKGPSTFGTPTAENPEYLGIDVP 1254
Qy      1261 V 1261
      1255 V 1255
Db

```

RESULT 10  
AAG88267  
ID AAG88267 standard; Protein: 1255 AA.

```

AC      AAG88267;
AC      11-SEP-2001 (first entry)
DT      11-SEP-2001 (first entry)
DE      HER2/neu amino acid sequence.
XX      Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
XX      immune response; vaccine; tumour; cancer; cytotoxic T lymphocyte; CTL.
XX      tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS      Homo sapiens.
XX      WO200141787-A1.
XX      14-JUN-2001.
XX      11-DEC-2000; 2000WO-US33591.
XX      10-DEC-1999; 99US-0458299.
XX      (EPIM-) EPIMMUNE INC.
XX      Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX      Keogh E;
XX      MPI; 2001-374995/39.

```

```

XX      An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT      cellular immune responses for the prevention and treatment of cancer -
XX
PS      Disclosure; Page 15; 1999p; English.
XX
XX      The present invention describes isolated prepared HER2/neu epitopes (1).
CC      Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC      culture in vitro and binds to a complex of an epitope (1), bound to a
CC      human leukocyte antigen (HLA) molecule; (2) a peptide (11) comprising (1)
CC      and a second epitope and the peptide is less than 50 contiguous amino
CC      acids that have 100% identity with a native peptide sequence of HER2/neu;
CC      (3) a vaccine composition (11) comprising (11) and a pharmaceutical
CC      excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC      (1); and (5) an isolated nucleic acid encoding (11). (11) has cytostatic
CC      and immunostimulant activities, and can be used in vaccines. (1), (11)
CC      and (11) are useful for inducing cellular immune responses for the
CC      prevention and treatment of cancer. (1) and (11) are useful for
CC      monitoring or evaluating an immune response to a tumour-associated
CC      antigen when incubated with a T lymphocyte sample from a patient and
CC      detecting the presence of bound T lymphocyte to (1) or (11). Epitope
CC      based vaccines mean that immunosuppressive epitopes that may be present
CC      in whole antigens may be avoided. Selected epitopes may be combined to
CC      enhance immunogenicity. The possible pathological side effects caused by
CC      infectious agents or whole protein antigen is eliminated. The vaccine
CC      provides the ability to direct and focus an immune response to multiple
CC      selected antigens from the same pathogen. Epitope-based anti-tumour
CC      vaccines provides the opportunity to combine epitopes derived from
CC      multiple tumour-associated molecules addressing the problem of tumour-
CC      tumour variability and reducing the likelihood of tumour escape due to
CC      antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC      the exemplification of the present invention.
XX
SQ      Sequence 1255 AA;

```

Query Match 96.8%; Score 6626; DB 22; Length 1255;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;

```

Qy      1 MELAALCRWGLLLALLPPGAASVQCTGDMKRLPASPEHLDMLRHLVGGCQVQGNL 60
Db      1 MELAALCRWGLLLALLPPGAASVQCTGDMKRLPASPEHLDMLRHLVGGCQVQGNL 60
Qy      61 ELTYLPTNASLSPFQDIQEVGVYLIANOVQVPLQRLIVRGOLFEDNYALAVDNG 120
Db      61 ELTYLPTNASLSPFQDIQEVGVYLIANOVQVPLQRLIVRGOLFEDNYALAVDNG 120
Qy      121 DPLNNTTPVTGASPGGLREQLRLSTELKGVLIQRNPOLCYOPTILMKDIFHNQOLA 180
Db      121 DPLNNTTPVTGASPGGLREQLRLSTELKGVLIQRNPOLCYOPTILMKDIFHNQOLA 180
Qy      181 LTLIDNRSRACHPCSPWCKSRGWSSESSDQSLTRTVAGGACARCGPLPTDCHEQC 240
Db      181 LTLIDNRSRACHPCSPWCKSRGWSSESSDQSLTRTVAGGACARCGPLPTDCHEQC 240
Qy      241 AAGCTGKHSQCLACHFNHSGICELCPALVYNTDTFESMPNPEGYTTGASCVTACP 300
Db      241 AAGCTGKHSQCLACHFNHSGICELCPALVYNTDTFESMPNPEGYTTGASCVTACP 300
Qy      301 YNYLSTVGSCTVLCPLHNOEVTAEQTORCEKSKPCARVCYGLGMQYIRANSKFIGIT 360
Db      301 YNYLSTVGSCTVLCPLHNOEVTAEQTORCEKSKPCARVCYGLGMHLEVRAYTSAN 360
Qy      361 ELEFAGCKKIIFGSLAFIPESFDGPASNTAFLQEOLOVFETLEITGYLYISAMPDSL 420
Db      361 IOEFAGCKKIIFGSLAFIPESFDGPASNTAFLQEOLOVFETLEITGYLYISAMPDSL 420
Qy      421 DLSVFOVLQVIRGILLNGAYSLTLOGLISWGLRLSRLREGSGALAIHNRTHLCFVHTV 480
Db      421 DLSVFOVLQVIRGILLNGAYSLTLOGLISWGLRLSRLREGSGALAIHNRTHLCFVHTV 480
Qy      481 PMDLFRFNFTVSFMLRVPKVASHLBEQVGEGLACHQLCARGHCGPGTQCVNCSQF 540
Db      481 PMDLFRFNFTVSFMLRVPKVASHLBEQVGEGLACHQLCARGHCGPGTQCVNCSQF 540

```

```

Db      481 PMDOLFRNPHQALLHTANRP-----DECVGEGLACHQLCARGHCMGPGPTQCVNCSQF 534
Qy      541 LRGOECVEECRVLOGLPREYVNAARHCLPCHPECOPONGSVTCGPEADOCVACAHYKDDP 600
Db      535 LRGOECVEECRVLOGLPREYVNAARHCLPCHPECOPONGSVTCGPEADOCVACAHYKDDP 594
Qy      601 FCVACPCSGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDLDDKCCPAEORASPLTSI 660
Db      595 FCVACPCSGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDLDDKCCPAEORASPLTSI 654
Qy      661 VSAVVGILLVVVLGVVFGILLIKRROOKIKKTYMRLLQSTELVEPLTSGANPNQOMRI 720
Db      655 ISAVVGILLVVVLGVVFGILLIKRROOKIKKTYMRLLQSTELVEPLTSGANPNQOMRI 714
Qy      721 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKYLRBENTSPKANKELIDAYYM 780
Db      715 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKYLRBENTSPKANKELIDAYYM 774
Qy      781 AGVSPYVSRLLIGICTSTVOLVTOIMPYGCLLDHVRENRGRLSGODLLNMCMQIAKMS 840
Db      775 AGVSPYVSRLLIGICTSTVOLVTOIMPYGCLLDHVRENRGRLSGODLLNMCMQIAKMS 834
Qy      841 YLEDVRLVHRDLAARVNLVKSPPNHVKITDGLARLLDIDETEVHADGGKVPKIMALESI 900
Db      835 YLEDVRLVHRDLAARVNLVKSPPNHVKITDGLARLLDIDETEVHADGGKVPKIMALESI 894
Qy      901 LRRRTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPOPICTIDVYMI 960
Db      895 LRRRTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPOPICTIDVYMI 954
Qy      961 MYKCMIDSECRPRRELVSERSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1020
Db      955 MYKCMIDSECRPRRELVSERSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1014
Qy      1021 GDLVAAEYLVQOGFPCDPAPGAGMWHHRSSSTSGGDLTLGIEPSEEPARSP 1080
Db      1015 GDLVAAEYLVQOGFPCDPAPGAGMWHHRSSSTSGGDLTLGIEPSEEPARSP 1074
Qy      1081 LAPSEGASDVEDGLGMGAAGLOSLPTHDPSPLORYSEDPVPLPSTEDGVAPLTS 1140
Db      1075 LAPSEGASDVEDGLGMGAAGLOSLPTHDPSPLORYSEDPVPLPSTEDGVAPLTS 1134
Qy      1141 POPEYVNOQDVAPQPPSPREGPLPAARPGATLERAKTSLPGKNGVVKDVPFAGAVENP 1200
Db      1135 POPEYVNOQDVAPQPPSPREGPLPAARPGATLERAKTSLPGKNGVVKDVPFAGAVENP 1194
Qy      1201 EYLTQOGGAPOPHPPAPAFNDLYYWDODPBERGAPSTFKGTPTAENBEYLGDDVP 1260
Db      1195 EYLTQOGGAPOPHPPAPAFNDLYYWDODPBERGAPSTFKGTPTAENBEYLGDDVP 1254
Qy      1261 V 1261
Db      1255 V 1255

```

## RESULT 11

AAE24067 ID AAE24067 standard; Protein; 1255 AA.

AAE24067;

23-SEP-2002 (first entry)

Human Her-2 protein.

Human: Her-2; epidermal growth factor receptor 2; infection; cancer;

hyperproliferative disorder; prophylaxis; inflammation; antisense;

tumour; gene therapy; phosphothioate backbone.

Homc. sapiens.

MO200222636-A1.

```

PD      21-MAR-2002.
XX
XX      12-SEP-2001; 2001WO-US28572.
PF
XX      15-SEP-2000; 2000US-0663834.
PR
XX      (ISIS-) ISIS PHARM INC.
PA
XX      Bennett CF, Cowselet LM;
PI
XX      WPI; 2002-471192/50.
DR      N-PSDB; AAD38904.
XX
XX      Novel antisense oligonucleotide which modulates the expression of Human
PT      Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT      inflammation or to prevent infection in humans.
XX
XX      Example 13; Page 95-107; 116pp; English.
XX
XX      The invention relates to antisense compounds targeted to a nucleic
XX      acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX      that specifically hybridises with and inhibits the expression of Her2.
XX      Antisense compounds of the invention are used for treating diseases or
XX      conditions associated with Her2 such as hyperproliferative disorders
XX      e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX      neutral or cardiac cancer. They are also useful prophylactically e.g.
XX      to prevent or delay infection, inflammation and tumour formation. The
XX      invention is also used in gene therapy. The present sequence is human
XX      Her-2 protein.

```

SQ Sequence 1255 AA;

Query Match 96.8%; Score 6626; DB 23; Length 1255;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;

```

Qy      1 MEIALACRWGLLALLPQAASTQVCTGDMRLRPASPTHLDMLRHLVYQCCQVYQGNL 60
Db      1 MEIALACRWGLLALLPQAASTQVCTGDMRLRPASPTHLDMLRHLVYQCCQVYQGNL 60
Qy      61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVQVPLQRLIRIVGTOLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVQVPLQRLIRIVGTOLFEDNYALAVLDNG 120
Qy      121 DPLNNTTPVTGASPGGLRELQRLSTELLKGVLIQRNPQLCYODTILMKDIFHKNNQLA 180
Db      121 DPLNNTTPVTGASPGGLRELQRLSTELLKGVLIQRNPQLCYODTILMKDIFHKNNQLA 180
Qy      181 LTLIDTNSRACHPCSPMKCKSGRCESESDQSLTRTYCAGACGACRCKGPLPTDCHEQC 240
Db      181 LTLIDTNSRACHPCSPMKCKSGRCESESDQSLTRTYCAGACGACRCKGPLPTDCHEQC 240
Qy      241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTFGASCYTACP 300
Db      241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTFGASCYTACP 300
Qy      301 YNYLSTDVGSCTLVCPRLNQEVTADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPRLNQEVTADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
Qy      361 ELEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPOLOVFEETLEETGLYLSAMPDSL 420
Db      361 IOEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPOLOVFEETLEETGLYLSAMPDSL 420
Qy      421 DLSVFQNIQVIRGRILHNGAYSVLTLQIGISWLGRLSRLRELGSGALLIHNTHLFCVHTV 480
Db      421 DLSVFQNIQVIRGRILHNGAYSVLTLQIGISWLGRLSRLRELGSGALLIHNTHLFCVHTV 480
Qy      481 PMDOLFRNMFVSWLVRPVKSASHLEECVGEGLACHQLCARGHCMGPGPTQCVNCSQF 540
Db      481 PMDOLFRNMFVSWLVRPVKSASHLEECVGEGLACHQLCARGHCMGPGPTQCVNCSQF 534
Qy      541 LRGOECVEECRVLOGLPREYVNAARHCLPCHPECOPONGSVTCGPEADOCVACAHYKDDP 600

```

```

Db      535  LRGQCEVECRVLCGLPREYNAHNLCPCHPECPQONGSVTFGBADQCVCAHKKDP 594
Qy      601  FCVARCPGKVPKSLSYMPIMKFPDEGACQPCPINCSTCVLDKDGCAEORASPLTSI 660
Db      595  FCVARCPGKVPKSLSYMPIMKFPDEGACQPCPINCSTCVLDKDGCAEORASPLTSI 654
Qy      661  VSAVVGILLVVLGVVFGILLIKRROOKIRKYMRLQETELVEPLTSGCAMPNOQMRI 720
Db      655  ISAVVGILLVVLGVVFGILLIKRROOKIRKYMRLQETELVEPLTSGCAMPNOQMRI 714
Qy      721  LKETELRKVKYLGSAGFTVYKGIWIPQENYKIVAIKVARENTSPKANKELIEAYVM 780
Db      715  LKETELRKVKYLGSAGFTVYKGIWIPQENYKIVAIKVARENTSPKANKELIEAYVM 774
Qy      781  AGVSGPYVSRLLGICLTSTVQLVTLQJMPYGLLDHVRNKGRLSGODLLNMCQIAKMS 840
Db      775  AGVSGPYVSRLLGICLTSTVQLVTLQJMPYGLLDHVRNKGRLSGODLLNMCQIAKMS 834
Qy      841  YLEVDRLVHRDLAANNVVKSPNHVKTIDFGLARLLDIDETBYHADGKVPICKMALESI 900
Db      835  YLEVDRLVHRDLAANNVVKSPNHVKTIDFGLARLLDIDETBYHADGKVPICKMALESI 894
Qy      901  LRRRFTHOSDVWSYGVVWELMTFGAKPYDGI PAREIPDLKESGRLPQPICTIDVYMI 960
Db      895  LRRRFTHOSDVWSYGVVWELMTFGAKPYDGI PAREIPDLKESGRLPQPICTIDVYMI 954
Qy      961  MYKCMIDSECRPRERELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1020
Db      955  MYKCMIDSECRPRERELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1014
Qy      1021  GDLYDAEYLVPOQCFPCPDPAKGGVNHHRSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db      1015  GDLYDAEYLVPOQCFPCPDPAKGGVNHHRSSSTRSGGDLTLGLEPSEEEAPRSP 1074
Qy      1081  LAPSEGASDVFDGDLGKAAKGLQSLPTHDPSPLOXYSEDPVTLPSSTOGVYAPLICS 1140
Db      1075  LAPSEGASDVFDGDLGKAAKGLQSLPTHDPSPLOXYSEDPVTLPSSTOGVYAPLICS 1134
Qy      1141  POPEVYVNDVVRPOPPSPREGPLPAARPAAGATLEBAKTLSPKNGVNDVFAFGAVENP 1200
Db      1135  POPEVYVNDVVRPOPPSPREGPLPAARPAAGATLEBAKTLSPKNGVNDVFAFGAVENP 1194
Qy      1201  EYLTPOGGAAPQPHPPAFSAFNDLYWDDPPERGAPESTFKCTPTAENPEYIGLDVP 1260
Db      1195  EYLTPOGGAAPQPHPPAFSAFNDLYWDDPPERGAPESTFKCTPTAENPEYIGLDVP 1254
Qy      1261  V 1261
Db      1255  V 1255

RESULT 12
AAE20479
ID      AAE20479 standard; Protein: 1255 AA.
XX      AAE20479;
XX      01-JUL-2002 (first entry)
XX      Human Her-2/neu protein.
XX      Human Her-2/neu protein.
XX      Human, Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX      human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX      Homo sapiens.
XX      Key
XX      Location/Qualifiers
XX      FT      1021..1030
XX      /note= "Naturally processed HLA-B*44-restricted epitope"
XX      MO200214503-A2.

```

```

PD      21-FEB-2002.
XX      XX
XX      14-AUG-2001; 2001WO-US41733.
XX      XX
PR      14-AUG-2000; 2000US-225152P.
XX      XX
PR      28-SEP-2000; 2000US-236428P.
XX      XX
PR      21-FEB-2001; 2001US-270520P.
XX      XX
XX      (CORI-) CORIXA CORP.
XX      XX
PI      Hand-Zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI      McNeill PD, Vedralck TS;
XX      XX
XX      WPI; 2002-280758/32.
XX      XX
XX      N-PSDB; AAD32743.
XX      XX
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX      prevention and diagnosis of cancer, preferably breast cancer
XX      XX
XX      Disclosure; Page 114-117; 129pp; English.
XX      XX
XX      The invention relates to an isolated Her-2/Neu polypeptide composition
XX      effective for eliciting an immune response. The invention is useful for
XX      eliciting an immune response in a patient, where the patient is human
XX      leukocyte antigen (HLA)-B*44 positive or is affected with breast cancer.
XX      The composition is useful for the therapy and diagnosis of cancer,
XX      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX      and other compositions for the diagnosis, prevention and treatment of
XX      human malignancies, for stimulating and/or expanding T cells specific for
XX      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX      patient. The invention is useful for stimulating a T cell response in a
XX      human patient, as probe or primer for nucleic acid hybridisation, to
XX      selectively form duplex molecules with complementary stretches of the
XX      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX      length gene from a suitable library, and to direct expression of a
XX      polypeptide in appropriate host cells. The composition is useful in
XX      prophylactic or therapeutic applications and for the treatment of cancer,
XX      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX      associated malignancies. The invention is useful in gene therapy. The
XX      present sequence is human Her-2/neu protein.
XX      XX
XX      Sequence 1255 AA:
XX      XX
XX      Query Match      96.8%; Score 6626; DB 23; Length 1255;
XX      Best Local Similarity 97.1%; Pred. No. 0;
XX      Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
XX      XX
Qy      1  MELAALCRWGLLALIPGAASVQCTGDMKLRLPASPEHLDMRLHYQCCOVQGNL 60
Db      1  MELAALCRWGLLALIPGAASVQCTGDMKLRLPASPEHLDMRLHYQCCOVQGNL 60
Qy      61  ELYLPTNASLSPLODIOEVGVYLIANNVROVPLQRLRYGRQLFEDNVALAVLDNG 120
Db      61  ELYLPTNASLSPLODIOEVGVYLIANNVROVPLQRLRYGRQLFEDNVALAVLDNG 120
Qy      121  DPLNNTPTVGTASPGELREQLRSLEILKGGVLIQRNPOLCYODTIIMKIDFHNKOLA 180
Db      121  DPLNNTPTVGTASPGELREQLRSLEILKGGVLIQRNPOLCYODTIIMKIDFHNKOLA 180
Qy      181  LTLIDITNRSRACHPCSPKCKSRGESSSEDCQSITRTVACAGCARCKGRLPTDCCHQC 240
Db      181  LTLIDITNRSRACHPCSPKCKSRGESSSEDCQSITRTVACAGCARCKGRLPTDCCHQC 240
Qy      241  AAGCTGPRHSQCLALHNHSGICELHCPALVYNTDTFESMPNPEGYTTGASCVTACP 300
Db      241  AAGCTGPRHSQCLALHNHSGICELHCPALVYNTDTFESMPNPEGYTTGASCVTACP 300
Qy      301  YNVLSTVGSCTLVCPHNOEVTADGTORCEKSKPCARVCYGLAMQYIRANSKFIGIT 360
Db      301  YNVLSTVGSCTLVCPHNOEVTADGTORCEKSKPCARVCYGLAMQYIRANSKFIGIT 360
Qy      361  ELEFAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVFETLEITGYLISAMPDSL 420
Db      361  ELEFAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVFETLEITGYLISAMPDSL 420

```

[illegible]

KM	Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KX	tyrosine kinase; receptor; c-erbB2; gene therapy.
OS	Homo sapiens.
FH	Key Location/Qualifiers
FT	Domain 1..653 /note= "extracellular domain"
FT	Domain 676..1255 /note= "intracellular domain"
FT	Domain 990..1255 /note= "phosphorylation domain"
PN	M0200212341-A2.
PD	14-FEB-2002.
PF	03-AUG-2001; 2001WO-US24283.
PR	03-AUG-2000; 2000US-0632507.
PA	(CORI-) CORIXA CORP.
XX	(SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
PI	Cheever MA, Gheyssen D;
DR	MP1_2002-241743/29.
DR	N-PsDB; ABA92250.
PT	Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT	or enhancing an immune response to the protein, has Her-2/neu
PT	extracellular domain fused to Her-2/neu intracellular or
PT	phosphorylation domain -
PS	Claim 68; Fig 7; 14tp; English.
XX	The present sequence is that of human Her-2/neu (p185 glycoprotein
CC	or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC	vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC	in a variety of cancers, including breast, ovarian, colon, lung and
CC	prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC	family of receptor-like glycoproteins. It comprises an extracellular
CC	domain with homology to the epidermal growth factor receptor
CC	(EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC	intracellular domain that also shows homology to EGFR. Its
CC	overexpression correlates with a poor prognosis in breast and
CC	ovarian cancers. The invention provides Her-2/neu fusion
CC	proteins, nucleic acids encoding them, viral vectors, and vaccines
CC	comprising the fusion proteins or nucleic acid molecules. In
CC	preferred fusion proteins, the extracellular domain of a Her-2/neu
CC	protein is fused to a Her-2/neu intracellular domain or
CC	phosphorylation domain (or its DeltaCD fragment). An immune
CC	response to Her-2/neu protein is elicited or enhanced by
CC	administering the fusion protein in the form of a vaccine, or by
CC	transfecting cells of an animal ex vivo with a nucleic acid
CC	encoding the fusion protein, and delivering the transfected cells
CC	to the animal. The fusion proteins, nucleic acids, and isolated
CC	specific T-cells are useful for inhibiting the development of a
CC	cancer, especially breast, ovarian, colon, lung or prostate cancer
CC	in a patient. T cells that specifically react with a Her-2/neu
CC	fusion protein can be used to remove tumour cells from a sample in
CC	order to inhibit the development of cancer in a patient.
XX	
SQ	Sequence 1255 AA:
Query Match	96.8%; Score 6626; DB 23; Length 1255;
Best Local Similarity	97.1%; Pred. No. 0;
Matches 1224; Conservative	9; Mismatches 22; Indels 6; Gaps 1.
QY	1 MELAAALRWGILLALLPRGAASVTCCTDVKLRLPASPETHLMLRRLYGCCGVQGNL 60
DB	1 MELDAARWGGILLALLPRGAASVTCCTDVKLRLPASPETHLMLRRLYGCCGVQGNL 60

```

QY 61 ELTLPNTNASTSLFLODIOEVGVYVLIANQVQVPLQRLRIYRGTOLEFEDNYALAVLNDG 120
DB 61 ELTLPNTNASTSLFLODIOEVGVYVLIANQVQVPLQRLRIYRGTOLEFEDNYALAVLNDG 120
QY 121 DPLNNTTAVTGSFPGSLRELOLRSITLTKGGVLIQRNPOLCYODTILMKDIFHNKOLA 180
DB 121 DPLNNTTAVTGSFPGSLRELOLRSITLTKGGVLIQRNPOLCYODTILMKDIFHNKOLA 180
QY 181 LTLIDTNSRACHPSPCKSGRCWSESESDCSLTJTVACAGCARCKGPLETDCCHQC 240
DB 181 LTLIDTNSRACHPSPCKSGRCWSESESDCSLTJTVACAGCARCKGPLETDCCHQC 240
QY 241 AAGCTGPGHSDCLALHNHNSGICELHCPALVYNTTDTFESPNPNEGRTYFASCVTACP 300
DB 241 AAGCTGPGHSDCLALHNHNSGICELHCPALVYNTTDTFESPNPNEGRTYFASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPFLHNOEVTADGTORCEKSKPCARVYCGIMOVYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPFLHNOEVTADGTORCEKSKPCARVYCGIMOVYIKANSKFIGIT 360
QY 361 ELLEFAGCKKIFGSLAFLEPSFDGDPASNTAPLQPELOVFEITLBEITGYLIYAMPDILP 420
DB 361 IOEFAGCKKITFGSLAFLEPSFDGDPASNTAPLQPELOVFEITLBEITGYLIYAMPDILP 420
QY 421 DLSVFQNIQVIRGRILHNGAYSLTQIGISWLGRLSIRELGSGLALIHNTHLCFVHTV 480
DB 421 DLSVFQNIQVIRGRILHNGAYSLTQIGISWLGRLSIRELGSGLALIHNTHLCFVHTV 480
QY 481 PMDOLFRNNFTVSWLHVPKVSASHLECEGEGJACHQOLCARGCMWGPPTQCNCSOF 540
DB 481 PMDOLFRNNFTVSWLHVPKVSASHLECEGEGJACHQOLCARGCMWGPPTQCNCSOF 540
QY 541 LRGECEVECRVLOGLPREYVNAHRLCPHECOPONSVTCTCFEADQCVACAHYKDP 600
DB 541 LRGECEVECRVLOGLPREYVNAHRLCPHECOPONSVTCTCFEADQCVACAHYKDP 600
QY 601 FCVARCPSGVKPDLSYMPIWKFPPDEEGACORCPINCHTSCVDLDKXCPAEGRASPLTST 660
DB 601 FCVARCPSGVKPDLSYMPIWKFPPDEEGACORCPINCHTSCVDLDKXCPAEGRASPLTST 660
QY 661 VSAVVGLILVVVLGVVFGILIKRROQIRKTYMRLLQETELVEPLTSGAMPNOAORI 720
DB 661 VSAVVGLILVVVLGVVFGILIKRROQIRKTYMRLLQETELVEPLTSGAMPNOAORI 720
QY 721 LKETELRKVVLGSGAFGTYYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILDEAVYM 780
DB 721 LKETELRKVVLGSGAFGTYYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILDEAVYM 780
QY 781 AGVSPVYSRLIGICTSTVOLVQOLMRYGCLLDHVRNRRGLSGODLLNMCQOIAKMS 840
DB 781 AGVSPVYSRLIGICTSTVOLVQOLMRYGCLLDHVRNRRGLSGODLLNMCQOIAKMS 840
QY 841 YLEEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDDIETEVHADGKVKIKMMALESI 900
DB 841 YLEEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDDIETEVHADGKVKIKMMALESI 900
QY 901 LRRRFTQSDVWSYGVTVMLMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDVYMI 960
DB 901 LRRRFTQSDVWSYGVTVMLMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDVYMI 960
QY 961 MVKCMWIDSECRPRFRRELVSFESRMARDPQRFVYIQNEDLPASPLDSTFRSLIEDDM 1020
DB 961 MVKCMWIDSECRPRFRRELVSFESRMARDPQRFVYIQNEDLPASPLDSTFRSLIEDDM 1020
QY 1021 GDVLDAEEYLVPOQGFPCPDPAFAGAGVHHRHSSSTRSGGDLTGLLEPSEEBAPSP 1080
DB 1021 GDVLDAEEYLVPOQGFPCPDPAFAGAGVHHRHSSSTRSGGDLTGLLEPSEEBAPSP 1080
QY 1081 LAPSEGASDVFDGDLGKAAGKLSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTC 1140
DB 1081 LAPSEGASDVFDGDLGKAAGKLSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTC 1140
QY 1141 POPEYVNOVDVRRPQPPSPREBPLPAARPAATLERAKTSLSPGKGVVQKDVAFGAVENP 1200
DB 1141 POPEYVNOVDVRRPQPPSPREBPLPAARPAATLERAKTSLSPGKGVVQKDVAFGAVENP 1200

```

```

DB 1135 POPEYVNOVDVRRPQPPSPREBPLPAARPAATLERAKTSLSPGKGVVQKDVAFGAVENP 1194
QY 1201 EYLTPOGGAAPQPPHPPAFSPAFDNLYYWDOPPRGAPSPSTFGKTPTAENEPEYGLDVP 1260
DB 1195 EYLTPOGGAAPQPPHPPAFSPAFDNLYYWDOPPRGAPSPSTFGKTPTAENEPEYGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
XX AAU77114;
XX
XX 05-JUN-2002 (first entry)
XX
XX Human Her-2/neu polypeptide.
XX
XX Human, Her-2/neu; cytostatic; haematological malignancy; CML;
XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
XX chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
XX Hodgkin's lymphoma; T cell therapy.
XX
XX Homo sapiens.
XX
XX WO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
XX encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide.
XX
XX Disclosure; Page 71-74; 74pp; English.
XX
XX The invention relates to a method for inhibiting development of
XX haematological malignancy in a patient by administering a polypeptide
XX comprising an immunogenic portion of Her-2/neu or a polynucleotide
XX encoding the polypeptide. Antigen presenting cells that express the
XX protein can also be administered. The sequences are used for inhibiting
XX development of haematological malignancy such as acute myelogenous
XX leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
XX leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
XX lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
XX Sequence 1255 AA;
XX
XX Query Match 96.8%; Score 6626; DB 23; Length 1255;
XX Best Local Similarity 97.1%; Pred. No. 0;
XX Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
QY 1 METALALRWKGLLALLPRGAASVCTGTMDKRLRPSPTHMLMRLHYGCGVQGNL 60
DB 1 METALALRWKGLLALLPRGAASVCTGTMDKRLRPSPTHMLMRLHYGCGVQGNL 60
QY 61 ELTLPNTNASTSLFLODIOEVGVYVLIANQVQVPLQRLRIYRGTOLEFEDNYALAVLNDG 120

```

```

Db      61 ELTYLPTNASLSFLDIOIEVQGVYLIAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy      121 DPLNNTTPTGTASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNQOLA 180
Db      121 DPLNNTTPTGTASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNQOLA 180
Qy      181 LTLIDTNSRACHPCSPMKCKSRCKGSESSDCCSLTRTVCAAGGCARCKGPLPTDCHEQC 240
Db      181 LTLIDTNSRACHPCSPMKCKSRCKGSESSDCCSLTRTVCAAGGCARCKGPLPTDCHEQC 240
Qy      241 AAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTFESNPNEGRTYFGASCVTACP 300
Db      241 AAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTFESNPNEGRTYFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCKCKPCARVCYGMQVITKANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCKCKPCARVCYGMQVITKANSKFIGIT 360
Qy      361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVFETLEITGYLISAMPDSL 420
Db      361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVFETLEITGYLISAMPDSL 420
Qy      421 DLSVTQNLQVTRIGRLIHNGAYSLTLOGLISWLGLRSLRELSGLALIHNTHLCPVHTV 480
Db      421 DLSVTQNLQVTRIGRLIHNGAYSLTLOGLISWLGLRSLRELSGLALIHNTHLCPVHTV 480
Qy      481 PMDOLFRFNFTVSMRLRVPKVSASHLECEVGEGLAQHQLCARGCMGSPQCVNCSQF 540
Db      481 PMDOLFRFNFTVSMRLRVPKVSASHLECEVGEGLAQHQLCARGCMGSPQCVNCSQF 540
Qy      541 LRGOCEVEECRLQGLPREYVNAHQLPCHPECOPONGSVTCFGEADQCVACAHYKDP 600
Db      541 LRGOCEVEECRLQGLPREYVNAHQLPCHPECOPONGSVTCFGEADQCVACAHYKDP 600
Qy      595 FCVACPSGVKPDLSYMPIMKFPDEBGAQPCPINCTHSCVDLDKCGCAEGRASPLTSI 654
Db      595 FCVACPSGVKPDLSYMPIMKFPDEBGAQPCPINCTHSCVDLDKCGCAEGRASPLTSI 654
Qy      661 VSAVVGILLVVLGVVFGILIKRROOKIRKYTRRLLOTELEVEPLTPSGAMPNOAKRI 720
Db      661 VSAVVGILLVVLGVVFGILIKRROOKIRKYTRRLLOTELEVEPLTPSGAMPNOAKRI 720
Qy      714 ISAVVGILLVVLGVVFGILIKRROOKIRKYTRRLLOTELEVEPLTPSGAMPNOAKRI 714
Db      714 ISAVVGILLVVLGVVFGILIKRROOKIRKYTRRLLOTELEVEPLTPSGAMPNOAKRI 714
Qy      721 LKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 780
Db      721 LKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 780
Qy      775 LKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 774
Db      775 LKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 774
Qy      781 AGVSPYVSRLLGICLTSTVQLVTQLMPYGCILLDVHRENRGLSGODLLNMCQIAKGS 840
Db      781 AGVSPYVSRLLGICLTSTVQLVTQLMPYGCILLDVHRENRGLSGODLLNMCQIAKGS 840
Qy      834 AGVSPYVSRLLGICLTSTVQLVTQLMPYGCILLDVHRENRGLSGODLLNMCQIAKGS 834
Db      834 AGVSPYVSRLLGICLTSTVQLVTQLMPYGCILLDVHRENRGLSGODLLNMCQIAKGS 834
Qy      841 YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGKVPKMMALSI 900
Db      841 YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGKVPKMMALSI 900
Qy      894 YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGKVPKMMALSI 894
Db      894 YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGKVPKMMALSI 894
Qy      901 LRRRTTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIDPLEKRGRLPQPICTIDVYMI 960
Db      901 LRRRTTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIDPLEKRGRLPQPICTIDVYMI 960
Qy      954 LRRRTTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIDPLEKRGRLPQPICTIDVYMI 954
Db      954 LRRRTTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIDPLEKRGRLPQPICTIDVYMI 954
Qy      961 MVKCMIMTISECRPRRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSFFYSLLLEDDM 1020
Db      961 MVKCMIMTISECRPRRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSFFYSLLLEDDM 1020
Qy      955 MVKCMIMTISECRPRRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSFFYSLLLEDDM 1014
Db      955 MVKCMIMTISECRPRRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSFFYSLLLEDDM 1014
Qy      1021 GDVDAEAYLVPOQGFPCPDPAFGAGVWHHRSSSTRSGGDLTLGLEPSEEAAPSP 1080
Db      1021 GDVDAEAYLVPOQGFPCPDPAFGAGVWHHRSSSTRSGGDLTLGLEPSEEAAPSP 1080
Qy      1074 GDVDAEAYLVPOQGFPCPDPAFGAGVWHHRSSSTRSGGDLTLGLEPSEEAAPSP 1074
Db      1074 GDVDAEAYLVPOQGFPCPDPAFGAGVWHHRSSSTRSGGDLTLGLEPSEEAAPSP 1074
Qy      1081 LAPSEGASDVFDGDLGMAAKGLOSLEPTHDSPLQRYSEDPVLPSSETDGYVALTCS 1140
Db      1081 LAPSEGASDVFDGDLGMAAKGLOSLEPTHDSPLQRYSEDPVLPSSETDGYVALTCS 1140
Qy      1075 LAPSEGASDVFDGDLGMAAKGLOSLEPTHDSPLQRYSEDPVLPSSETDGYVALTCS 1134
Db      1075 LAPSEGASDVFDGDLGMAAKGLOSLEPTHDSPLQRYSEDPVLPSSETDGYVALTCS 1134
Qy      1141 POPEYVNOVDYRPOPPSPREGLPARAPAGATLEBAKTLPGKNVVDVFAFGAVENP 1200
Db      1141 POPEYVNOVDYRPOPPSPREGLPARAPAGATLEBAKTLPGKNVVDVFAFGAVENP 1200

```

```

Db      1135 POPEYVNOVDYRPOPPSPREGLPARAPAGATLEBAKTLPGKNVVDVFAFGAVENP 1194
Qy      1201 EYLTPQGAAPQPHPPAPAFNDLYYWDODPERGAPPSPTFKGTPTAENPEYGLDVP 1260
Db      1195 EYLTPQGAAPQPHPPAPAFNDLYYWDODPERGAPPSPTFKGTPTAENPEYGLDVP 1254
Qy      1261 V 1261
Db      1255 V 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
AA39568;
AC AAR39568;
AD 07-FEB-1994 (first entry)
DE Sequence of c-erbB-2 tumour antigen.
DM Tumour antigen; c-erbB-2; glycoprotein.
OS Homo sapiens.
PN WO9316185-A.
PI 19-AUG-1993.
PD 05-FEB-1993; 93WO-US01055.
PF 06-FEB-1992; 92US-0831967.
PR (CETU ) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Houston TX, Oppermann H, Ring DB;
DR MPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
PS Disclosure; pages 48-54; 87pp; English.
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells. Such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083; AAR39568). X in AAR39568 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 96.2%; Score 6583; DB 14; Length 1433;
Best Local Similarity 96.5%; Pred. No. 0;
Matches 1217; Conservative 11; Mismatches 27; Indels 6; Gaps 1;

Qy      1 MELAALCEWGLLALIPGAASTOVCTGDMKRLPASPTHLDMLRHLVQGCQVVGNTL 60
Db      1 MELAALCEWGLLALIPGAASTOVCTGDMKRLPASPTHLDMLRHLVQGCQVVGNTL 60
Qy      61 ELTYLPTNASLSFLDIOIEVQGVYLIAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASLSFLDIOIEVQGVYLIAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy      121 DPLNNTTPTGTASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNQOLA 180
Db      121 DPLNNTTPTGTASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNQOLA 180
Qy      181 LTLIDTNSRACHPCSPMKCKSRCKGSESSDCCSLTRTVCAAGGCARCKGPLPTDCHEQC 240
Db      181 LTLIDTNSRACHPCSPMKCKSRCKGSESSDCCSLTRTVCAAGGCARCKGPLPTDCHEQC 240
Qy      1261 LTLIDTNSRACHPCSPMKCKSRCKGSESSDCCSLTRTVCAAGGCARCKGPLPTDCHEQC 240
Db      1261 LTLIDTNSRACHPCSPMKCKSRCKGSESSDCCSLTRTVCAAGGCARCKGPLPTDCHEQC 240

```

QY 241 AAGCTGPHGSDCLACLFHFNHSGICELHCPALVTYNTDTFESPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPHGSDCLACLFHFNHSGICELHCPALVTYNTDTFESPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCCTLVCPLEHNOEVTADGTORCEKSKPCARVCYGLMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCCTLVCPLEHNOEVTADGTORCEKSKPCARVCYGLMEHLREVAVTSAN 360  
QY 361 ELEFAGCKKIFGSLAFLEPSEFDGDPASNTAPLQPEOLQVFETLEETGLYISAMPDSL 420  
Db 361 IOEFAGCKKIFGSLAFLEPSEFDGDPASNTAPLQPEHLQVFETLEQITGLYISAMPDSL 420  
QY 421 DLSVFNQIVIRGRILHNGAVSLTQGLGISWLGRLSRLRELSGLALIHNTLCEVHTV 480  
Db 421 DLSVFNQIVIRGRILHNGAVSLTQGLGISWLGRLSRLRELSGLALIHNTLSFVHTV 480  
QY 481 PMDOLFRNNFTVSWLVRPKVSASHLECEVGEGLACHOLCARGHCWGPPTQCVNCSOF 540  
Db 481 PMDOLFRNPFOALHTANRPE----DECVGEGLACHOLCARGHCWGPPTQCVNCSOF 534  
QY 541 LRGOECVECEVLOGLPREYVNAHQLPCHPECOPONGSVTCFGEADQCVACAHYKDP 600  
Db 535 LRGOECVECEVLOGLPREYVNAHQLPCHPECOPONGSVTCFGEADQCVACAHYKDP 594  
QY 601 FCVABCPGCVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCCPAEQRASPLTSI 660  
Db 595 FCVABCPGCVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCCPAEQRASPLTSI 654  
QY 661 USAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOAMRI 720  
Db 655 ISAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOAMRI 714  
QY 721 LKETELRKVKYLGSGAFGTYYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780  
Db 715 LKETELRKVKYLGSGAFGTYYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 774  
QY 781 AGVSPYVSRLLGICLTSTVOLVTOAMPYGLLDHVRENRGLSQDILLNMCQIAKMS 840  
Db 775 AGVSPYVSRLLGICLTSTVOLVTOAMPYGLLDHVRENRGLSQDILLNMCQIAKMS 834  
QY 841 YLEDEVRLVHRDLAARNVLYKSPNHVKITDPGLARLLDIDETEHADGGKVPKMMALESI 900  
Db 835 YLEDEVRLVHRDLAARNVLYKSPNHVKITDPGLARLLDIDETEHADGGKVPKMMALESI 894  
QY 901 LRRRFTHQSDVMSYGVTVMEMLTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960  
Db 895 LRRRFTHQSDVMSYGVTVMEMLTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954  
QY 961 MVKCMIMIDSECRPRPRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1020  
Db 955 MVKCMIMIDSECRPRPRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1014  
QY 1021 GDLVDABEYLVPOGGFFCPDPAFGAGVHHRHSSSTRSGGDLTLGLEPSEEEA PRSP 1080  
Db 1015 GDLVDABEYLVPOGGFFCPDPAFGAGVHHRHSSSTRSGGDLTLGLEPSEEEA PRSP 1074  
QY 1081 LAPSEGAGSDVFDGDLGWAAGKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1140  
Db 1075 LAPSEGAGSDVFDGDLGWAAGKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1134  
QY 1141 POPEYVNOQDVVRPOPSBREGPLPAARPAGATLERAKTTLSPGKGVVVDVAFGAVENP 1200  
Db 1135 POPEYVNOQDVVRPOPSBREGPLPAARPAGATLERAKTTLSPGKGVVVDVAFGAVENP 1194  
QY 1201 EYLTPOGGAAPQPPAPAFSPFNDLYWMDPPERGA PPFKGTPTAENPEYLGLDVP 1260  
Db 1195 EYLTPOGGAAPQPPAPAFSPFNDLYWMDPPERGA PPFKGTPTAENPEYLGLDVP 1254  
QY 1261 V 1261  
Db 1255 V 1255

Search completed: July 22, 2003, 08:41:40  
Jcb time : 44.1589 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using SW model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds  
(without alignments)  
5267.077 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 6841  
Sequence: 1 MELALCRMGLLALLPPGA.....TFKGTPTANPEYLGLDVPV 1261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6621	96.8	1255	1 A24571	protein-tyrosine k
2	5827	85.2	1260	1 TVRINU	protein-tyrosine k
3	5817.5	85.0	1254	2 I48161	p-185 precursor -
4	3110	45.5	1210	1 GOUHE	epidermal growth f
5	3079	45.0	1210	2 A53183	epidermal growth f
6	3053.5	44.6	1223	1 TVCHLV	epidermal growth f
7	2945.5	43.1	1308	2 A47253	epidermal growth f
8	2627	38.4	1166	1 S06142	epidermal growth f
9	2378.5	34.8	1342	2 A36223	kinase-related tra
10	2293.5	33.5	1339	2 JCA387	epidermal growth f
11	1725.5	25.2	698	1 TVPVLV	protein-tyrosine k
12	1677	24.5	604	1 TVTTH	protein-tyrosine k
13	1625.5	23.8	1330	1 GQFE	epidermal growth k
14	1606	23.5	544	2 S35745	epidermal growth f
15	1599	23.4	545	2 S00727	protein-tyrosine k
16	1582	23.1	540	2 B44776	kinase-related tra
17	1580	23.1	540	1 TVPVEB	protein-tyrosine k
18	1464	21.4	644	2 A36225	epidermal growth k
19	1305	19.1	1123	2 E88257	epidermal growth f
20	1305	19.1	1374	2 S70712	protein-tyrosine k
21	1220	17.8	1369	2 S70713	protein-tyrosine k
22	1176	17.2	1717	1 A45558	epidermal growth f
23	1126	16.5	527	2 A42032	epidermal growth f
24	979.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	730	10.7	1363	2 T43220	insulin-like growth
28	696	10.2	1382	1 INHUR	insulin receptor p
29	694	10.1	1383	2 A36080	insulin receptor p

30	693	10.1	1477	2 T18534	protein-tyrosine k
31	691.5	10.1	1372	2 A34157	insulin receptor p
32	691	10.1	1607	2 T43212	insulin-like growth
33	687.5	10.0	1300	2 A36502	insulin receptor-x
34	678	9.9	1268	2 B36502	insulin receptor-x
35	647	9.5	1367	1 IGHURI	insulin-like growth
36	628	9.2	1371	2 A33837	insulin-like growth
37	620.5	9.1	2148	1 A56081	insulin receptor -
38	620	9.1	1330	2 T30346	insulin receptor -
39	614	9.0	2101	2 S57245	insulin receptor (
40	597	8.7	987	2 A54092	protein-tyrosine k
41	594.5	8.7	977	2 S49004	tyrosine kinase Mp
42	591.5	8.6	984	2 A39753	protein-tyrosine k
43	588.5	8.6	976	2 A36355	protein-tyrosine k
44	586.5	8.6	952	2 I50612	protein-tyrosine k
45	586	8.6	1091	2 S33596	protein-tyrosine k

## ALIGNMENTS

### RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein erl

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.

Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human C-erb-B-2 gene to epidermal growth

A/Reference number: A24571; PMID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epide:

A/Reference number: A25491; PMID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,

Science 230, 1137-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chrom

A/Reference number: A44188; PMID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU1>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

A/Residues: 1-517, 'RALV', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; PMID:85272597; PMID:2992089

A/Accession: I59509

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:L29395; NID:G459607; PIDN:AAA35809.1; PID:G459808

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlesinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcription;

A/Reference number: I57622; PMID:87286998; PMID:3039351

A/Accession: I57622

A/Molecule type: DNA

A/Residues: 1-191 <YAL>

A:Cross-references: GB:M16792; NID:g183983; PIDN:AA58637.1; PID:g553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 3d  
 C:Genetics: GDB:ERBB2; NGL; NEU; HER-2  
 A:Gene: GDB:ERBB2; NGL; NEU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: the list of introns is incomplete  
 C:Function: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-155/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:22-653/Domain: extracellular #status predicted <EXT>  
 F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:654-675/Domain: transmembrane #status predicted <TM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:68,124,187,259,530,571,629/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:753/Active site: Lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.8%; Score 6621; DB 1; Length 1255;  
 Best Local Similarity 96.8%; Pred. No. 2.6e-267;  
 Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

Qy	1	MEALALCRMGILLALPPGAAS	TOVCTGTPMKLR	LPASPETHLMRLHLYOGC	VOVGNL	60				
Db	1	MEALALCRMGILLALPPGAAS	TOVCTGTPMKLR	LPASPETHLMRLHLYOGC	VOVGNL	60				
Qy	61	ELTYLPTNASL	IFLODIQEVGVY	LHNOVQVPLQRL	IVRGSTQ	120				
Db	61	ELTYLPTNASL	IFLODIQEVGVY	LHNOVQVPLQRL	IVRGSTQ	120				
Qy	121	DPLNTPPVY	GASGGLREQLRSL	TEILKGVLIQNNPOL	CTYODTILMDIFHKNQ	180				
Db	121	DPLNTPPVY	GASGGLREQLRSL	TEILKGVLIQNNPOL	CTYODTILMDIFHKNQ	180				
Qy	181	LTLIDTRSR	ACHPCSPMCKGSR	CMGSESDCOSL	TRTVCA	240				
Db	181	LTLIDTRSR	ACHPCSPMCKGSR	CMGSESDCOSL	TRTVCA	240				
Qy	241	AAGCTGPK	HSDCCLALCFHNSG	ICELHCPALVTYNTD	FESMPN	300				
Db	241	AAGCTGPK	HSDCCLALCFHNSG	ICELHCPALVTYNTD	FESMPN	300				
Qy	301	YNTLSTVGS	CTLVCPILHNOE	VTAEDEGTQREKSK	CARVCYGLGMQY	1KANKF	360			
Db	301	YNTLSTVGS	CTLVCPILHNOE	VTAEDEGTQREKSK	CARVCYGLGMQY	1KANKF	360			
Qy	361	ELFPA	GCKKIFGSLAF	PESFGDPASNTAP	LOPEOLVFETLEIT	GYLYISAMP	420			
Db	361	ELFPA	GCKKIFGSLAF	PESFGDPASNTAP	LOPEOLVFETLEIT	GYLYISAMP	420			
Qy	421	DLVSFONL	QVIRGRILHNGAV	SLTLQGLISWLG	ILRSLSREL	SGSLAIHNTHL	480			
Db	421	DLVSFONL	QVIRGRILHNGAV	SLTLQGLISWLG	ILRSLSREL	SGSLAIHNTHL	480			
Qy	481	PMOOLP	FNPHOALHTANR	PEDECVGGLACHOU	CAAGH	CMGSRPTOCVNC	540			
Db	481	PMOOLP	FNPHOALHTANR	PEDECVGGLACHOU	CAAGH	CMGSRPTOCVNC	540			
Qy	541	VEECRV	LVQGLPREYVNA	RHCLPCH	PECO	PONGSVTCG	PEADQCAVA	600		
Db	541	VEECRV	LVQGLPREYVNA	RHCLPCH	PECO	PONGSVTCG	PEADQCAVA	600		
Qy	601	PFNNFTV	SPFLRVKVSASH	LE---	PDEB	ACOPCP	INTCHSCVD	LDKGC	PAEOR	656
Db	601	PFNNFTV	SPFLRVKVSASH	LE---	PDEB	ACOPCP	INTCHSCVD	LDKGC	PAEOR	656

Db	601	PSG-----	VKPDLSYMP	IKWPDEB	GACOPCP	INTCHSCVD	LDKGC	PAEOR	650	
Qy	657	LTSIVSA	VAVGILLVVLGV	VPFGILIKRQOK	IKIRKTYTRR	LLQELTE	VELPE	LTPSG	AMP	716
Db	651	LTSIVSA	VAVGILLVVLGV	VPFGILIKRQOK	IKIRKTYTRR	LLQELTE	VELPE	LTPSG	AMP	710
Qy	717	QMR	LKETE	LRKVVLG	SGAFGT	YKGIWIPD	GENVXI	PVAIK	YLR	776
Db	711	QMR	LKETE	LRKVVLG	SGAFGT	YKGIWIPD	GENVXI	PVAIK	YLR	770
Qy	777	AYWAG	SVYSRLG	ICLSTVQ	VTQMLPY	QCLLDH	VENR	GR	LSQ	836
Db	771	AYWAG	SVYSRLG	ICLSTVQ	VTQMLPY	QCLLDH	VENR	GR	LSQ	830
Qy	837	KGMSY	LEDVLR	YRDLA	RVLYKS	PNHVKIT	DPGLAR	LLDID	TE	896
Db	831	KGMSY	LEDVLR	YRDLA	RVLYKS	PNHVKIT	DPGLAR	LLDID	TE	890
Qy	897	LESLRR	RFTHOS	DVMSY	GYTVME	LMFGAK	PKYGI	PAR	ELPDL	956
Db	891	LESLRR	RFTHOS	DVMSY	GYTVME	LMFGAK	PKYGI	PAR	ELPDL	950
Qy	957	VYIMV	KWMI	DSECR	PRFELV	SEFS	RMARD	PFV	IONED	1016
Db	951	VYIMV	KWMI	DSECR	PRFELV	SEFS	RMARD	PFV	IONED	1010
Qy	1017	DDMDG	LDVA	BEVYL	VPQGF	FCDDP	PA	GAG	GVHRR	1076
Db	1011	DDMDG	LDVA	BEVYL	VPQGF	FCDDP	PA	GAG	GVHRR	1070
Qy	1077	PRSP	LAP	EGAS	DPVFGD	LGMA	KGLQSLPT	HDSP	LO	1136
Db	1071	PRSP	LAP	EGAS	DPVFGD	LGMA	KGLQSLPT	HDSP	LO	1130
Qy	1137	LTCSP	QPE	EVYNQ	DPVR	QPPSP	PREGL	PAAR	PAGAT	1196
Db	1131	LTCSP	QPE	EVYNQ	DPVR	QPPSP	PREGL	PAAR	PAGAT	1190
Qy	1197	VENPE	YLT	PPGGA	PPHPP	PAF	PAF	PAF	PAF	1256
Db	1191	VENPE	YLT	PPGGA	PPHPP	PAF	PAF	PAF	PAF	1250
Qy	1257	LDV	VPV	1261						
Db	1251	LDV	VPV	1255						

RESULT 2  
 TVRTNU  
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Dec-1988 #sequence-revision 31-Dec-1988 #text-change 11-Jun-1999  
 C:Accession: A24562; A61204  
 R:Barbmann, C.I.; Hung, M.C.; Weinberg, R.A.  
 Nature 319, 226-230, 1986  
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.  
 A:Reference number: A24562; MUID:86118662; PMID:3945311  
 A:Accession: A24562  
 A:Molecule type: mRNA  
 A:Residues: 1-1260 <BAR>  
 R:Mani, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen  
 Carcinogenesis 12, 1975-1978, 1991  
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no  
 2-thiathiolylformamide or N-methyl-N-nitrosourea.  
 A:Reference number: A61204; MUID:92035293; PMID:1682063  
 A:Accession: A61204  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 637-663, 'V', 665-702 <MAS>  
 A>Note: authors translated the codon GCA for residue 25 as Val  
 C:Genetics:  
 A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAM>  
 F:658-680/Domain: transmembrane #status predicted <TMN>  
 F:723-988/Domain: protein kinase homology <KIN>  
 F:731-739/Region: protein kinase ATP-binding motif  
 F:711,791,263,535,576,634/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:758/Active site: Lys #status predicted  
 F:882,1227/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.3%; Score 5827; DB 1; Length 1260;  
 Best Local Similarity 85.3%; Pred. No. 2e-234;  
 Matches 1081; Conservative 53; Mismatches 117; Indels 16; Gaps 4;

```

Oy 1 MELAALCRMGILALALPPGAASTOVCTGDMKRLPASPEITHLDMRLHYOGCQVQGNL 60
Db 4 MELAAMCRMGILALALPPGAGTQVCTGDMKRLPASPEITHLDMRLHYOGCQVQGNL 63
Oy 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOQVPLQRLRIYRGTOQLFEDNYALAVLDNG 120
Db 64 ELTYVPANASLSFLQDIOEVQGVYLAHNOQVPLQRLRIYRGTOQLFEDNYALAVLDNR 123
Oy 121 DPLANTTPVT-GASPGGLRELQRLSTELIKGVLQORNQOLCYODTIIMKDIIFHKNNOL 179
Db 124 DPDNNVAASPTGRTPEGLRELQRLSTELIKGVLIRGNPOLCYODMVLKMDVFRKNNOL 183
Oy 180 ALTLIDTNSRACHPCSPMKSGRSGESSEDCOSLTRITACAGGACRCKGLPTDCCHEQ 239
Db 184 APLDIDTNSRACHPCSPACKNDHNCWGESPEDCOILGITITGSGACRCKRLPTDCCHEQ 243
Oy 240 CAAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDFESNPNEGRTYFGASCVTAC 299
Db 244 CAAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDFESNPNEGRTYFGASCVTTC 303
Oy 300 PNYLSTDVGSCTLVCPRLNQEYTAEDGTORCKCKSPCARVCYIGIAMOYIKANSRFIGI 359
Db 304 PNYLSTEVGSCTLVCPRLNQEYTAEDGTORCKCKSPCARVCYIGIAMETHLRGARAITS 363
Oy 360 TELEFAGCKKIFCSLFLPESPDGPASNTAPLOPOLQVFEETLGTGLYISAMPDLS 419
Db 364 NVOEPGCKKIFCSLFLPESPDGPASNTAPLOPOLQVFEETLGTGLYISAMPDLS 423
Oy 420 PDLVSFQNLQVIRGRILHNGAYSLSLQGLISWLGRLSRELGSGLALIHNNHLCFVHT 479
Db 424 RDLVSFQNLRIIRGRILHNGAYSLSLQGLISWLGRLSRELGSGLALIHNNHLCFVHT 483
Oy 480 VMDQLFRPHQALLHTANRPEDE-CVGGSLACHOLCANGHCCKGPPPTCCVNCSGFLRQ 538
Db 484 VMDQLFRPHQALLHTANRPEDE-CVGGSLACHOLCANGHCCKGPPPTCCVNCSGFLRQ 543
Oy 539 ECVEECRVLQGLPREVYNARHCLPCHPEGQPNQSGVTCFGRPADQCVACAHYDPFCVA 598
Db 544 ECVEECRVLQGLPREVYNARHCLPCHPEGQPNQSGVTCFGRPADQCVACAHYDPFCVA 603
Oy 599 RCPFNNTVSPMLRVKVSASHLE---PDEGACQPCPINCINCHSCVLDKCKCPAEPORA 654
Db 604 RCPFSG-----VKPDLSTYMPKIKYDDEGICQPCPINCINCHSCVLDKCKCPAEPORA 653
Oy 655 SPLTISVAVNGILLVNVGVVFGILLIKRQOKIRKTYRRLLOETELVEPLTPSGAMPN 714
Db 654 SPVTIATIAVEVLLFLILVVVVGILLIKRQOKIRKTYRRLLOETELVEPLTPSGAMPN 713
Oy 715 QAKMRLKETELRKVYVLSGAFGVYKGIWI PDGENVNI PVAIKYLRNTSPKANKETL 774
Db 714 QAKMRLKETELRKVYVLSGAFGVYKGIWI PDGENVNI PVAIKYLRNTSPKANKETL 773
Oy 775 DEAYVAVGVSPPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYRENGRLGSDOLNMCMQ 834
Db 774 DEAYVAVGVSPPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYRENGRLGSDOLNMCMQ 833
Oy 835 IAKGMSYLEDVLRHDLAARNVLYKSPNHVKITPDGLARLLDIDETETHADGKVPYIKW 894

```

```

Db 834 IAKGMSYLEDVLRHDLAARNVLYKSPNHVKITPDGLARLLDIDETETHADGKVPYIKW 893
Oy 895 MALESTILRRRFTQSDVMSYGVVWMLMTFGAPYDGI PARPITDLEKGRLLPOPICT 954
Db 894 MALESTILRRRFTQSDVMSYGVVWMLMTFGAPYDGI PARPITDLEKGRLLPOPICT 953
Oy 955 IDVYIMVWKCMIDSECRPRFRELVEFSFMAADPQRFVYIYONEDLGPASPLDSTFYRSL 1014
Db 954 IDVYIMVWKCMIDSECRPRFRELVEFSFMAADPQRFVYIYONEDLGPASPLDSTFYRSL 1013
Oy 1015 LEDDDMGDLVDAEELVLPQCGFCPPDPAAGGVHHRHRSSTRSGGDLTLGLEPSEE 1074
Db 1014 LEDDDMGDLVDAEELVLPQCGFCPPDPAAGGVHHRHRSSTRSGGDLTLGLEPSEE 1073
Oy 1075 EAPRSPAPSEAGSDVPDDDLGMGAAGKGLPTHDPSLQVRSSEPTPLPSETGYV 1134
Db 1074 GPRSPAPSEAGSDVPDDDLGMGAAGKGLPTHDPSLQVRSSEPTPLPSETGYV 1133
Oy 1135 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAAPAGATLEBRATLSPGKGVVADVAFG 1194
Db 1134 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAAPAGATLEBRATLSPGKGVVADVAFG 1193
Oy 1195 GAVENBEYLVTPQCGAALPQHPHPPAFSPAFDNLVYWDPPERGAIPSTFGTPTAENPEY 1254
Db 1194 GAVENBEYLVTPQCGAALPQHPHPPAFSPAFDNLVYWDPPERGAIPSTFGTPTAENPEY 1253
Oy 1255 LGLDVPV 1261
Db 1254 LGLDVPV 1260

```

RESULT 3  
 148161  
 p-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_rev1500 02-Jul-1996 #text\_change 18-Jun-1999  
 C:Accession: 148161  
 R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,  
 Gene 140, 251-255, 1994  
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Reference number: 148161; M0ID:94193007; PMID:7908275  
 A:Accession: 148161  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <RES>  
 A:Cross-references: GB:D16295; NID:9493236; PIDN:BA03801.1; PID:9747595  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:718-983/Domain: protein kinase ATP-binding motif  
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.0%; Score 5817.5; DB 2; Length 1254;  
 Best Local Similarity 85.0%; Pred. No. 5e-234;  
 Matches 1075; Conservative 62; Mismatches 113; Indels 15; Gaps 3;

```

Oy 1 MELAALCRMGILALALPPGAASTOVCTGDMKRLPASPEITHLDMRLHYOGCQVQGNL 60
Db 1 MELAAMCRMGILALALPPGAGTQVCTGDMKRLPASPEITHLDMRLHYOGCQVQGNL 60
Oy 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOQVPLQRLRIYRGTOQLFEDNYALAVLDNG 120
Db 61 ELTYVPANASLSFLQDIOEVQGVYLAHNOQVPLQRLRIYRGTOQLFEDNYALAVLDNR 120
Oy 121 DPLANTTPVTGASPGGLRELQRLSTELIKGVLQORNQOLCYODTIIMKDIIFHKNNOLA 180
Db 121 DPLANTTPVTGASPGGLRELQRLSTELIKGVLQORNQOLCYODTIIMKDIIFHKNNOLA 180
Oy 122 DPLDNTTATGRTPEGLRELQRLSTELIKGVLIRGNPOLCYODTIIMKDIIFHKNNOLA 180
Db 122 DPLDNTTATGRTPEGLRELQRLSTELIKGVLIRGNPOLCYODTIIMKDIIFHKNNOLA 180
Oy 181 LTLIDTNSRACHPCSPMKSGRSGESSEDCOSLTRITACAGGACRCKGLPTDCCHEQ 240
Db 181 PVDIDTNSRACHPCSPACKNDHNCWGESPEDCOILGITITGSIAPRAVAPARARLLPTDCCHEQ 240

```

```

QY 241 AAGCTGKSHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYFGASCYACP 300
DB 241 AAGCTGKSHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYFGASCYACP 300
QY 301 YNYLSTVSGCTTLCPLHNOEVTAEADGTORCEKSKPCARCYGCMQYIKANSKEFIGIT 360
DB 301 YNYLSTVSGCTTLCPLHNOEVTAEADGTORCEKSKPCARCYGCMQYIKANSKEFIGIT 360
QY 361 ELEPAGCKKIFGSLAFIPESFDGDPASNTAPLQPOLOVFETLEITGYLYISAMPDLP 420
DB 361 IQEPAQCKKIFGSLAFIPESFDGDPASNTAPLQPOLOVFETLEITGYLYISAMPDLP 420
QY 421 DLSEFOQLQVIRGRILNNGAYSLTLOGIGLSMLRSRLRELSGLALIHNTHLCPVHTV 480
DB 421 DLSEFOQLQVIRGRILNNGAYSLTLOGIGLSMLRSRLRELSGLALIHNTHLCPVHTV 480
QY 481 PMDQLFNNPQALHTANRPEDEVCSEGLACHOLCARGHCWGPPTQCVNCSOFLRQEC 540
DB 481 PMDQLFNNPQALHTANRPEDEVCSEGLACHOLCARGHCWGPPTQCVNCSOFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHEPCOPONGSVTCGPEADQCAHYKDPPEVCARC 600
DB 541 VKECRVWKGLPREYVNGKHCCLPCHPECCOPONSTETCTGSEADQCTACPHYKDPPEVCARC 600
QY 601 PFNNFTVSFMLRVPKVSASHLE----PDEGACQPCPINCNCSCVDLDDKCPAEORASP 656
DB 601 PSC-----VKPDLSTWPIWKYPRBESMCPCLNCHSCVDLDERCPAEORASP 650
QY 657 LTSIVSAVGLLVVLGVVFGILIKRQOKIRKYTRKRLQETLEVEPLTPSGAMPNOA 716
DB 651 ATSIIVAVGILFLVIGVVGILIKRRQKIRKYTRKRLQETLEVEPLTPSGAMPNOA 710
QY 717 QMRILKTELRKXVAVLSSGAGTYKGIWIPDGENVKIPVAKYLRENTSKANKELIDE 776
DB 711 QMRILKTELRKXVAVLSSGAGTYKGIWIPDGENVKIPVAKYLRENTSKANKELIDE 770
QY 777 AYWAGSGPYVSVLLGLCTSTVQOLVTQOLMPYGCCLDHYENGRGLSQOOLMWCQOIA 836
DB 771 AYWAGSGPYVSVLLGLCTSTVQOLVTQOLMPYGCCLDHYENGRGLSQOOLMWCQOIA 830
QY 837 KGSYLEDVRLVHRDLAARNVLVKS PMNVKITDGLARLLDIDETEVHADGKVPYIKMA 896
DB 831 KGSYLEDVRLVHRDLAARNVLVKS PMNVKITDGLARLLDIDETEVHADGKVPYIKMA 890
QY 897 LESLIRRRFTHQSVMSVGYVWELMTFGARPYGCIIPAREIPDLLEKBERLPPPICTID 956
DB 891 LESLIRRRFTHQSVMSVGYVWELMTFGARPYGCIIPAREIPDLLEKBERLPPPICTID 950
QY 957 VYMIWVCMWIDSECRPRELVESEFSRMARDPQRFVYIQMEDLGSPASPLDSTFYRSLLE 1016
DB 951 VYMIWVCMWIDSECRPRELVESEFSRMARDPQRFVYIQMEDLGSPASPLDSTFYRSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFCCPDPAQAGGMVHRRHSSSTRSGGDLTLGLEPSEBEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDPAQAGGMVHRRHSSSTRSGGDLTLGLEPSEBEA 1070
QY 1077 PRSLASBEGAGSVPGDLCMGAKGLQSLPTHDPSPLOQYSSDPYVPLPSEFDGYAP 1136
DB 1071 PRSLASBEGAGSVPGDLCMGAKGLQSLPTHDPSPLOQYSSDPYVPLPSEFDGYAP 1130
QY 1137 LTCSPOPEYVNPQVPRPQPSPRREGPLPAARPAAGATLERAKTSLPGKNGVYQVFAFGA 1196
DB 1131 LACSPQPEYVNPQVPRPQPSPRREGPLPAARPAAGATLERAKTSLPGKNGVYQVFAFGA 1190
QY 1197 VENPEYTLPOGGAAPQHPPPAPSPAFDNLVYMDODPERGAPSTKGTPTAENPEYL 1256
DB 1191 VENPEYLVPRGSGASQPH-PPALCPAFDNLVYMDODPERGSGSPNTEGPTAENPEYL 1249
QY 1257 LDVPU 1261
DB 1250 LDVPU 1254

```

## RESULT 4

## GQHE

Epidermal growth factor receptor precursor - human

N/Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C/Species: Homo sapiens (man)

C/Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999

C/Accession: A00641, A25772, S30024, A38672, A0642, A3365, A23062, A05281, A60143, A

R/Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.

rg, P.H.

Nature 309, 418-425, 1984

A/Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression

A/Reference number: A00641, MUID:84219729, PMID:6328312

A/Accession: A00641

A/Molecule type: mRNA

A/Residues: 1-1210 <URL>

A/Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA52240.1; PID:g757924

A/Note: the authors translated the codon AAG for residue 540 as Asn

Rishi, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.

Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985

A/Title: Characterization and sequence of the promoter region of the human epidermal g

A/Reference number: A25772, MUID:85270438, PMID:2991899

A/Accession: A25772

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-29 <ISH>

A/Cross-references: GB:M11234; NID:g181991; PIDN:AAA52370.1; PID:g553272

R/Haley, J.; Whittle, N.; Bennett, P.; Kitchington, D.; Ullrich, A.; Waterfield, M.

Oncogene Res. 1, 375-396, 1987

A/Title: The human EGF receptor gene: structure of the 110 kb locus and identification

A/Reference number: S30024, MUID:8821333, PMID:3329716

A/Accession: S30024

A/Molecule type: DNA

A/Residues: 1-29 <HA2>

A/Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119

R/Haley, J.D.; Waterfield, M.D.

J. Biol. Chem. 266, 1746-1753, 1991

A/Title: Contributory effects of de Novo transcription and premature transcript termin

A/Reference number: A38672, MUID:91107677, PMID:1988448

A/Accession: A38672

A/Molecule type: DNA

A/Residues: 1-29 <HL2>

A/Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271

A/Experimental source: carcinoma cell line A431-7

R/Xu, Y.; Ishii, S.; Clark, A.O.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; M

Nature 309, 806-810, 1984

A/Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNA

A/Reference number: A00642, MUID:84245835, PMID:6330563

A/Accession: A00642

A/Molecule type: mRNA

A/Residues: RCAMRRA, 150-187, 'KSVIQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-3

, 798-799, 'TD', 802-811, 'R', 813-942 <XUV>

A/Experimental source: A431 human carcinoma cells, which have large numbers of EGF rec

R/Lin, C.R.; Chen, W.S.; Krutiger, W.; Stolarsky, L.S.; Weber, M.; Evans, R.M.; Verma,

Science 224, 843-848, 1984

A/Title: Expression cloning of human EGF receptor complementary DNA: gene amplificatio

A/Reference number: A43615, MUID:84196372, PMID:6326261

A/Accession: A43615

A/Molecule type: mRNA

A/Residues: 713-964 <LIN>

A/Experimental source: epidermoid carcinoma cell line A431

R/Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.

Biochem. Biophys. Res. Commun. 124, 125-132, 1984

A/Reference number: A23062, MUID:85046483, PMID:6093780

A/Accession: A23062

A/Molecule type: mRNA

A/Residues: 1028-1210 <SIM>

R/Weber, M.; Gill, G.N.; Speiser, J.

Science 224, 294-297, 1984

A/Reference number: A05281, MUID:84172183, PMID:6324343

A/Accession: A05281

A/Molecule type: protein

A/Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R.Russo, M.W.; Lukae, T.J.; Cohen, S.; Stares, J.V.  
 J. Biol. Chem. 260, 5205-5208, 1985  
 A/Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor  
 A/Reference number: A60143; MUID:85182650; PMID:2985580  
 A/Accession: A60143  
 A/Molecule type: protein  
 A/Residues: 740-744, X, 746-747 <RUS>  
 R:Motczkowski, B.; Mosig, G.; Cohen, S.  
 Nature 309, 270-273, 1984  
 A/Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase  
 A/Reference number: A38023; MUID:84191554; PMID:6335948  
 A/Contents: annotation, receptor activity  
 A/Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.; Cell 59, 33-43, 1989  
 A/Title: Functional independence of the epidermal growth factor receptor from a domain A:Reference number: A3331; MUID:9000233; PMID:2790960  
 A/Contents: annotation, internalization signal  
 C/Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex  
 C/Genetics:  
 A:Gene: EGFR  
 A:Cross-references: GDB:120610; OMIM:131550  
 A:Map position: 7p12.3-7p12.1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase; signal sequence #status predicted <SIG>  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-1210/Product: EGF receptor #status predicted <MAT>  
 F:25-645/Domain: extracellular #status predicted <EXT>  
 F:75-300/Domain: EGF receptor extracellular domain repeat <EET>  
 F:130-600/Domain: EGF receptor extracellular domain repeat <EET>  
 F:446-668/Domain: transmembrane #status predicted <TM>  
 F:669-1210/Domain: intracellular #status predicted <INT>  
 F:710-975/Domain: protein kinase homology <KIN>  
 F:718-726/Region: protein kinase ATP-binding motif  
 F:999-1046/Region: coated-pit mediated internalization signal  
 F:1047-1210/Region: inhibitory  
 F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:745/Active site: Lys #status experimental

Query Match 45.54; Score 3110; DB 1; Length 1210;  
 Best local Similarity 49.64; Pred. No. 7, 6e-122; Indels 124; Gaps 26;  
 Matches 633; Conservative 170; Mismatches 350;

```

QY 11 LLLALLPPGAA--STOVCTGTDMKRLPASPTHLDMLRYOCQVVOGNETLYLPTN 68
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 14 LLAALCPASRLREKVKCGQTSNKLTLQGFEDHPLSLQRMFNNECEVLSNLETTYQRN 73
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 69 ASLSFLODIOVGVLIANNQVROVPLQRLVIRGTQLFEDNYALAVLDNGPLNNTTP 128
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 74 YDLSFLKTIQEVGVYLIANTVERIPLNLIQIRGMVYENSVALAVLSNYD----- 126
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 129 VTGASFGGLRELDRLSITELIKGVLIQRPOLCYODTILMKDIFHNQNLATLLIDTR 168
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 127 ---ANKTGLKELPMRLQELIHGAVRPSNNPALCNVESIQMRDVISSDFLSNMKDFQNH 183
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 189 SPACHPGSPWCKSGSRGSSSEDQSLTRTVCAAGCA-RCKGRLPTDCHEQCAAGCTGP 247
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 184 LGSQCKDSCSPGSCWGAEEENCOKLTKIIQAQCSGRCGRKSPSCCHNQCAAGCTGP 243
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 248 KMSDCLACHFNHSGICELHCPALVYNTDFTESMRPERRYFGASCCTACPYNTLSTD 307
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 244 RSDSDLVCKRFRDEATCKTCTPPLMLYNTPTTYMDVNPBEKVSFGATCVKCCPRNYVTD 303
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 308 VSSCTLVCPHNOEVAEDGTORCEKSPKCAVCGVGLGMQYIKANSKFIETLE-FAG 366
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 304 HSGSCVACGADSDTEM-EEDGVKRCCKCEGCRKVCIGIGIGERK-DLSLNATNTHKFKN 361
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 367 CKKIFGSLAFLPSPGDPASNTAPLQPEGLQVFTLEETIGLYTISAMPDLSPLDSVFG 426
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 362 CTSISGDLHLPAFAFGDSFTTHPPPLDPOELDKTKVETITGFLLIQAMPEKNTDLHAFE 421
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 427 NLOVRGRILNHGAVSLTQGLGISWLGSLRELSSGLALHNHNLCAVTHVPDOLF 486
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

DB 422 NLEIIRRTKQHGQPSLAIVSLNITSLSGLSLKEISDGVIIISGNKMLCYANTINMKKLF 481
QY 487 RNPQALLHTNRPDEDECVGGLACHOLCARGHCWGPPTQCVNCSQPLRGCEVEECRV 546
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 482 GTSQKTKIISNRKENSCKATQVCHALCSPEGCGEPBPBCVCSNRVSRGECVDCKL 541
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 547 LOGLPREYVNAHRLCPHRECOPOGNSVTCGPEPADQCAVAKYDPFCVACRCP----- 602
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 542 LEGPREFEVESECIQCHPECLPQAMNITCTGRBPNDICQAHIDGPHCVKTCGAVWG 601
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 603 -NNTVFWLVRPVKVSASHLEPDEEGACOPRPNICTSHSCVDLDDKCAEGRASPLTSIV 661
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 602 ENNTLV--W--KYADAGH-----CHLCPNCTGCTGPGEGCPTNGPKIP--SIA 647
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 662 SAAYG---ILLVVLGVVFGILIRROOKIRKYMRLQETTELVEPTSPGAMNOQM 718
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 648 TGMVGLALLLVVALGIG--LFRKRHHYAKRTLRLQRELVEPLTPSGEARNQML 704
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 719 RLKETEELRKVKVLSGAFGTVYKGIWIPDENYKIPVAIKVRENTSPKANKETLDEAY 778
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 705 RLKETEELRKVKVLSGAFGTVYKGIWIPDENYKIPVAIKVRENTSPKANKETLDEAY 764
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 779 VMAGVSPYVSLIGLITLSTVQLVTQMLPYGCLLDHYRENRGLSGODLLNMCQIAKG 838
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 765 VMASVDNPHVCRLLGICLTSTVQLITQIMPFGCLLDYREKDNIGSQVLLNMCVQIAKG 824
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 839 MSYLEDRLVHRLDAARNVLKVSNNHYKINDPGLARLLIDETETHADGCVPIKMALE 898
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 825 KMYLEDRLVHRLDAARNVLKVPQHVKITDPGLAKLGAEEKYVHAGGVPIKMALE 884
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 899 SILRRRTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVY 958
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 885 SILHRIYTHOSDWSYGVTVWELMTFGSKPYDGIPIAREIPDLLEKGERLPQPICTIDVY 944
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 959 MIMVCKMIDSECRPRRELIVSEFSRNARDPQRFVITQ-NEDLCPASPLDSTFYRSLSD 1017
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 945 MIMVCKMIDADSPKFERELIIEFSKVARDPQRLVITQGBERHMLPSPTSNFFRYALMDE 1004
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1018 DDMGDLVDAEBEYLVPQCGFCPPDPAFAGCMVHHRHSSSTRSGGDLTLGFESEEAR 1077
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1005 EDMDDVDVDADEYLLPQCGFF-----SSPETS 1030
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1078 RSLPAPSEAGSDVDFDGLGMAAGLSLPTHDPSPLQRYSDPTVLPSET--DGVA 1135
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1031 RTPLSLSLSTASN--NSTVACIDRNGLOSCPIKEDSLQKYSPTGALTEDSIDDTFL- 1087
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1136 PLTCSQPEVYNQDVARPPSPREGRPLPARPAGATLERAKTILSPKNGVGVKDFAFGG 1195
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1088 -----PVPEYING-SVPKRPAQSVQNVHNPQPLNP-----APSRDPHYQD--PHST 1131
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1196 AVENPEYL--TPQGAAPQHPPPAPFAPDNLYTWDQ-----DP-----PERGA 1238
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1132 AVGNPEYLVQ-----PTCVNSTFDSPAHMAKGSHQISLDNPVQOQDFPFXAK 1162
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1239 PPSTFGKPTPAENPEYL 1255
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1183 PNGIFKGS-FAENAEYL 1198
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 5  
 A53183  
 Epidermal growth factor receptor precursor - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999  
 C/Accession: A53183; A43818; S24942; A28941; S45325; I49643  
 R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A. Genes Dev. 8, 399-413, 1994  
 A/Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor  
 A/Reference number: A53183; MUID:94170986; PMID:8125255  
 A/Accession: A53183  
 A/Molecule type: mRNA  
 A/Residues: 1-1210 <LUS>  
 A/Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
Oncogene 6, 673-676, 1991  
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site  
A:Reference number: A43818; MUID:91232866; PMID:2030916  
A:Accession: A43818  
A:Molecule type: mRNA  
A:Residues: 1-714 <AVI>  
A:Cross-references: GB:X59698  
R:Esinger, D.P.; Serrero, G.  
Submitted to the EMBL Data Library, June 1992  
A:Reference number: S24942  
A:Accession: S24942  
A:Molecule type: mRNA  
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>  
A:Cross-references: EMBL:Z12608  
R:Heisermann, G.U.; Gill, G.N.  
J. Biol. Chem. 263, 13152-13158, 1988  
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated  
A:Reference number: A28941; MUID:88310814; PMID:3138233  
A:Accession: A28941  
A:Molecule type: Protein  
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,  
R:Hubbs, M.L.; Dunn, A.R.; Alexander, W.S.  
Submitted to the EMBL Data Library, April 1994  
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor  
A:Reference number: S45325  
A:Accession: S45325  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-971, 'K', 973-1210 <VER>  
A:Cross-references: EMBL:X78987; MUID:9488830; PIDN:CAA5587.1; PID:9488831  
R:Paria, B.C.; Dae, S.K.; Andrews, G.K.; Dey, S.K.  
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993  
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse  
A:Reference number: 149643; MUID:93126380; PMID:7678348  
A:Accession: 149643  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 12-20, 22-132 <RES>  
A:Cross-references: GB:L06864; MUID:9193001; PIDN:AAA53029.1; PID:9567201  
C:Genetic8:  
A:Gene: EGFR  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:648-670/Domain: transmembrane #status predicted <TM>  
F:712-977/Domain: protein kinase homology <KIN>  
F:720-728/Region: protein kinase ATP-binding motif  
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental  
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental  
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental  
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental  
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.0%; Score 3079; DB 2; Length 1210;  
Best Local Similarity 49.5%; Pred. No. 1.5e-120;  
Matches 635; Conservative 161; Mismatches 360; Indels 128; Gaps 27;

Qy 11 LLLALLPPGAA--STVCTGTDMKLRIPASPEHILMDKRLHYOGCQVVGLELTPTN 68  
Db 14 LRLTLCAGAGALEEKVCOCSTSNRLTOLGTEDHFLSLORMYNNCEVGLTETTYVRN 73  
Qy 69 ASLSFLDIOEVGVVLIHNOVROVPLORLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128  
Db 74 YDLSFLKTIQEVAGVLIANTVERIPLENLIIRGNALYENTYALALSLN----- 124  
Qy 129 VTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYDITILWKDI-----FKRONLATLI 184  
Db 125 -YGNRTGLRELPLRNLOEILIGAVRFENNPILCNMIDTIQWRDIQVNVFMSMNDL--- 180  
Qy 185 DTRNSRACHPSCSPCKSGKRCWGESSEDCQSLTRVTCAGGCA-RCKGPLPTCCGECQCAAG 243  
Db 181--QSHPSSCPKCDPSCPNCSGWGGEENOCQLTKITICAOQCSHRCGRSPSCCHNQCAAG 239

Qy 244 CTGPKHSDCLACLHFHNSGICELHCPALVNTYNTDTPSPMNPREGRYTPGACVTAAPVNY 303  
Db 240 CTGRESDDCLVCQKFOEATCTKOTCPPLMLYNTPTTYQDVNPEGKTSFGATCVKCKPNY 299  
Qy 304 LSTDVGSCTVLCPLHNOEVTAEADGTORCEKSKPCARVCYGLQNOYIKANSKFTIGTELE 363  
Db 300 VYDHDGSCVACGPDYEV--BEDGIRCKKCKDDGCRVNCNGIGIGEFR-DLTSINATNIK 357  
Qy 364 -PACCKKIPGSLAFLPSPFDGDPASNTAPLOPELOVPELTLEITGYLYISAMVDSLPLD 422  
Db 358 HFXYCTAISGDLHLLPAFAFKGDSFTRTPLDPRELEIKTYKEITGFLLIQAMPDMNTDL 417  
Qy 423 SVPNLOVIRGRILHNAVSLTLOGLSIMWLSRLSELGSLALIHNNHLCVHTVPM 482  
Db 418 HAFENLEIRKTKOHOQFSLAVVGLNITSGLSLKLEISGDVYISGNRLCYANTIM 477  
Qy 483 DOLFNRHQAHLHTANPEDECEVEGALCHQLCARGCWGPGPYCVNCVSOLFQGECEVE 542  
Db 478 KKLFGTNOKTKIMNNAEAKDCAVNHVNCPLGSEGCWGPEDCVSCQVNSRGRECEVE 537  
Qy 543 ECRVLOGLPBYVNAHCLPCHPECPQNGSVTCFEPADQCAVCAHYKDPFCVACPF 602  
Db 538 KCNILEGPEPEFVNSSECIOCHPECLPQAMNITCTGGRPDNCAHYIDGPHCVKTCPA 597  
Qy 603 ----NPFVSFWLRVPRKVSASHLEPDEBAGCQPCINCTHSQVLDLQKCPAEGRASPL 657  
Db 598 GIMGENNTLV--MKYA-----DANNVCHLCHACTYGCAGCPGLQGEVWPSEKXI 645  
Qy 658 TSIVSAVVGILLVVLGVVFGI-LIKRQOKIRKRYTRRLLOETLEVEPLTPSGAMPQOA 716  
Db 646 PSINTGVIGGLFLIV--VALGIGLPMRRHIVRKRTRRLLOEELVEPLTPSGAPQOA 704  
Qy 717 QMRILKTELRKRVKVLGSAFGTYKGIWIPDGNVKI PVAIKVLRNTSPKANKIIDE 776  
Db 705 HLRIKETEPEFKIIVLGSAGFGTYKGIWIPDEKVKI PVAIKVLRNTSPKANKIIDE 764  
Qy 777 AYVWAGSPVPSRLIGLSTVQLVTLQMPYGLLDHVENRNGRSGOOLLMMQOIA 836  
Db 765 AYVWASVDNPRVCHLGLCLSTVQLTQLMPYGLLDYVNEHKNIGSOYLLMNCVQIA 824  
Qy 837 KGNLYEDVRLVHRDLAARVLYKSPNNVKTITDFGLARLIDIDEYHADGKVPKIMA 896  
Db 825 KGNVYLEDRLVHRDLAARVLYKTPQHVKTITDFGLAKLGAEEKEYAEGKVPKIMA 884  
Qy 897 LESTLRRFTHQSVNSGYVTWELMTFGAKRYGICIPARETIPDLKEKERLPORPYCTID 956  
Db 885 LESTLHRIYTHQSVNSGYVTWELMTFGSKPYDGIIPASDISILEKERLPORPYCTID 944  
Qy 957 VYMTMVKCMIMDSRCRPRFRELVESEFRMARDPQRFVIO--NEDLGPAASPLDSTFYRSL 1015  
Db 945 VYMTMVKCMIMDASRPREFRLEIFESOMARDPQRYLVIOGDERMHLPSPTDSNRYRALM 1004  
Qy 1016 EDDMDGLVDAAEYLVYQOGFFCDDPAPAGAGMHHRRSSSTSGGDLTLGLEPSEE 1075  
Db 1005 DEEDMEDVDADEYLVYQOGFF-----NSST----- 1031  
Qy 1076 APRSPPLAPBSGAGDVFDGLGMAAGKLOSLPTHDSPLQRYSEDPVLPSET--DGY 1133  
Db 1032 -SRTPLLSLSATSN---NSTVACINRNGSCRVEDAFLORYSSDPGATVEDNIDDAF 1086  
Qy 1134 VAPLTGSPQPEYVQPVPRPSPRSGPLPAARPAAGTTERATLTSQKGVVVDVAF 1193  
Db 1087 L-----PVEEYVQ--SVPKRPAGSVQNPVHNPRLH-----ARGRLHYON--PH 1129  
Qy 1194 GGAVENTEYL--TPQGAAPQPPAPSPAPDNLVYMDQ-----DP-----PER 1236  
Db 1130 SNAVNGEYLVNTAQ-----PTCISGSGNSPALMTQKSHQMSLDNPNYQDDFFPKE 1180  
Qy 1237 GAPPSTFKGPTTANPEYLGIDVP 1260  
Db 1181 TKPNGIFKG--PTAEMAYLRLVAPP 1203